# Canada's premier challenge

*Justin Trudeau's Liberal Party has struggled to live up to its promises on the environment. Whoever wins the coming election can and must do better.*

Canada's Prime Minister Justin Trudeau — the young, liberal political leader — swept into power in 2015 with a decisive victory and an air of new hope. He promised to bring change to a country facing environmental havoc, 'muzzled' government scientists and disappointing science budgets. He made scientific evidence part of his brand, and pledged to elevate the status of science in government.

Four years on, Trudeau's government has lost its shine. It has kept many of its science promises, but has fallen short on key pledges, notably on the environment. Perhaps as a result, his Liberal Party is facing a much tougher reception among voters as he runs for re-election on 21 October.

As *Nature* went to press, the outcome of the election was too close to call: polls predict an even split between the right-wing conservatives and Trudeau's left-wing liberals, with the further-left New Democratic Party and the Green Party picking up the slack. Whoever wins, evidence-based policies and the environment must be at the heart of their agenda.

In its 2015 election campaign, the Liberal Party promised to right what it saw as many scientific and environmental wrongs, and in government it made several high-profile wins. The new cabinet included a dedicated minister for science, Kirsty Duncan, who was, in turn, mandated to appoint a chief science adviser to the federal government — both striking changes from the previous government.

Then, during its first month of office, Trudeau's government declared federal scientists free to speak to the media, attempting to rectify the practice of permission-seeking that had been established under the conservative administration. Duncan went on to commission the first comprehensive review of the nation's science research structure in decades. The Fundamental Science Review called on the government to re-establish a strong footing for fundamental research.

These moves are welcome, but it is too soon to say whether the use of scientific evidence in decision-making has been strengthened — it took two years before molecular biologist Mona Nemer was picked for the post of chief science adviser.

One of Nemer's mandates was to continue to address the 'muzzling' issue. A survey in 2017 found that more than half of government scientists still felt they could not speak freely, but, under Nemer's guidance, a robust and much-lauded batch of Scientific Integrity Policies followed, making it clear that federal scientists can speak about their work without requiring approval. As for the Fundamental Science Review, its recommendations have yet to be fully taken on board by the government.

Trudeau's budgets, too, have flip-flopped between wins and disappointments for Canada's researchers. The first budget, in 2016, brought a windfall that doubled the funding boost for the main granting agencies compared with the previous year; but this was followed by a surprising flatline budget in 2017, which killed off Canada's Climate Change and Atmospheric Research programme.

The following year, scientists rallied to the cause of promoting science in the run-up to the budget, and 2018's funding was hailed as the largest investment in science in Canadian history; but, by contrast, the 2019 budget brought only small spending bumps. Several pressure groups and scientific societies have now banded together to make science a core election issue and highlight what remains to be done.

Perhaps most disappointing of all has been the Trudeau government's inability to live up to its environmental promises in the face of economic pressures.

On 17 June 2019, the federal government declared a climate emergency, and last month Greta Thunberg helped to draw 500,000 people out onto Canada's streets in one of the largest ever environmental protests. But, at present, Canada is unlikely to meet its self-declared goal of reducing greenhouse-gas emissions to 30% below 2005 levels by 2030. The federal government's carbon tax policy is being challenged in some of the provinces — and has no support from conservatives — and a pledge to become carbon neutral by 2050 remains vague. Even a poll of the government's own scientists found that an overwhelming majority are dissatisfied with climate-change policies.

> *"A sustainable future backed by the best evidence needs leadership."*

On the bright side, according to the Global Cleantech Innovation Index (last updated in 2017), Canada rose from seventh out of 40 nations in 2014, to fourth in 2017.

In 2016, the government developed a Can$1.5-billion (US$1.1-billion) Oceans Protection Plan; and in 2018, it championed a non-binding Ocean Plastics Charter at the G7 meeting. That same year, Nemer also established an Independent Expert Panel on Aquaculture, to feed scientific evidence into policy decisions in what has proved to be a thorny area of conflict between industry and environmentalists.

But these achievements pale in comparison to investments in fossil fuels. The government controversially spent Can$4.5 billion buying an oil pipeline expansion project from the Kinder Morgan energy company, to ensure that more oil from Alberta's tar sands gets to the west coast for export. Trudeau recently promised that profits from the pipeline will be used to pay for a Can$3-billion fight against climate change, including the planting of 2 billion trees — 200 million a year, on top of the 600 million currently planted annually. But this has not mollified environmentalists or Indigenous rights activists, who are infuriated by the purchase.

The continuing challenge for whichever party wins this October will be to forge a leadership role for Canada in sustainable development — against the interests of the fossil-fuel industry. Canada ranks third in the world for proven oil reserves, thanks to its tar sands; that's a powerful economic force. But the country also has the talent and investment in new technologies — from artificial intelligence to quantum computing — to make it a global leader in an emerging economy more directed at a sustainable future.

A sustainable future backed by the best evidence needs leadership. It is a challenge that no prime minister can afford to shirk. ∎

# Gandhi on science

*The champion of India's freedom movement was a supporter of sustainable development.*

India's tourist shops do a good trade in Gandhi memorabilia. One particularly popular souvenir is a plaque that lists Mohandas Karamchand (Mahatma) Gandhi's 'seven social sins'. These include 'politics without principles', 'commerce without morality' and 'science without humanity'.

During his lifetime and after his assassination in January 1948, Gandhi, the human-rights barrister turned freedom campaigner, has been mischaracterized as anti-science — often because of his concerns over the human and environmental impacts of industrial technologies.

But in the month that the world commemorates the 150th anniversary of Gandhi's birth, it is time to revisit our understanding of this aspect of his life and work. Gandhi was a keen student of the art of experimentation — his autobiography is subtitled 'The Story of My Experiments with Truth'. He was an enthusiastic inventor and an assiduous innovator, making, discarding and refining snake-catching tools, sandals made from used tyres, and methods for rural sanitation, not to mention the small cotton-spinning wheels that would become his trademark.

Anil Gupta at the Indian Institute of Management in Ahmedabad, who has researched rural innovation in India for 40 years, says that Gandhi was also an early adopter of developing and improving technologies using crowd-sourcing — in 1929 he announced a competition, with a cash prize, to design a lightweight spinning wheel that could produce thread from raw cotton. It would be of solid build quality that would last for 20 years. "Gandhi was an engineer at heart," adds Anil Rajvanshi, director of the Nimbkar Agricultural Research Institute in Phaltan, India.

Gandhi adopted experimental methods equally in his planning and execution of civil-disobedience campaigns against colonial rule. That legacy alone has endured to the extent that climate-change protest groups such as Extinction Rebellion describe themselves as following in a Gandhian tradition.

Gandhi drew the line at the resource-intensive, industrial-scale engineering that Britain brought to India after the first waves of the Industrial Revolution. Inspired in part by the writings of Ralph Waldo Emerson, John Ruskin, Henry David Thoreau and Leo Tolstoy, he called for manufacturing on a more human scale, in which decisions about technologies rested with workers and communities.

Gandhi was aware that he was perceived as being anti-science. His biographer Ramachandra Guha quotes a 1925 speech to college students in Trivandrum (now Thiruvananthapuram) in southern India, in which Gandhi said that this misconception was a "common superstition". In the same address, he said that "we cannot live without science", but urged a form of accountability: "In my humble opinion there are limitations even to scientific search, and the limitations that I place upon scientific search are the limitations that humanity imposes upon us."

*"He was an enthusiastic inventor and an assiduous innovator."*

Gandhi understood that technology's negative impacts are often felt disproportionately by low-income rural populations. In that same speech to the Trivandrum students, he challenged his young audience to think of these communities in their work. "Unfortunately, we, who learn in colleges, forget that India lives in her villages and not in her towns. How will you infect the people of the villages with your scientific knowledge?" he asked them.

In the end, Gandhi's call for less-harmful technologies was out of sync with India's newly independent leadership, and also went against the grain of post-Second World War science and technology policy-making in most countries. India's first prime minister, Jawaharlal Nehru, was strongly influenced by European industrial technology and also by the model of large publicly funded laboratories — the forerunners to today's vibrant and globally renowned institutes of science and technology. By contrast, Gandhi's ideas were seen as quaint and impractical.

Influential figures from history often leave contested legacies. But in one respect at least, the space for debate about Gandhi's life and impact has narrowed. As the world continues to grapple with how to respond to climate change, biodiversity loss, persistent poverty, and poor health and nutrition, Gandhi's commitment to what we now call sustainability is perhaps more relevant today than in his own time. ∎

# Nile tensions

*Let researchers finish their work on the impacts of Africa's largest hydropower dam.*

Scientists investigating the hydrology of the Nile are likely to have heard the story of their tenth-century predecessor, mathematician and physicist Ibn al-Haytham. The ruler of Egypt asked al-Haytham to dam the river, but it proved too great an engineering challenge. Fearing the caliph's wrath, al-Haytham is said to have feigned illness to avoid being punished.

Thankfully, the scientists currently advising Egypt, Ethiopia and Sudan on the Grand Ethiopian Renaissance Dam do not face anything like the same risks. But they are nevertheless under pressure as talks between the three countries — and especially between Egypt and Ethiopia — have hit an impasse (see page 159).

Ethiopia says the hydropower dam is needed urgently, because two-thirds of the country has lacked electricity for too long. Egypt is in less of a hurry. Ninety per cent of its fresh water comes from the Nile, and it is concerned that the dam will create water scarcity for its 100 million inhabitants over the five to seven years needed to fill the dam's reservoir. Last week, Egypt decided that it wants another country to mediate the dispute — naming the United States as its preferred choice. Ethiopia rejects this proposal. This is an unfortunate turn of events. There might well be a need for mediation, but now is too soon. The countries are still waiting for the outcome of an independent scientific assessment of the dam's risks to downstream countries.

In 2015, Egypt, Ethiopia and Sudan agreed that an expert panel, the National Independent Scientific Research Group (NISRG), would assess the environmental impacts of each country's preferred timetable for constructing the dam. The group has been meeting regularly and is preparing to produce a consensus report and provide recommendations. But Egypt's decision to call for mediation before the scientists have had a chance to report puts the NISRG in an awkward position: the researchers representing Egypt, especially, might feel pressure not to write or say anything that could undermine their government's negotiating position.

Instead of rushing straight into mediation, the countries should let their scientific advisers complete the task that has been asked of them. The researchers should be allowed to publish their findings for scrutiny by everyone concerned, not least the citizens of the three countries, who will be most affected by the dam.

International involvement might be needed if the scientific advisers are unable to produce a consensus report, or if, once the findings are published, political leaders are unwilling or unable to shift their positions. But until then, Egypt, Ethiopia and Sudan need to let the researchers finish the job they have been asked to do. ∎

JERRI CALDWELL HAMMONDS

# Banish hunger on university campuses

*For World Food Day,* **Esther Ngumbi** *calls on institutions of higher education to help students know where their next meal is coming from.*

When I had to skip meals to pay for rent during my student days at Kenyatta University in Nairobi, Kenya, studying became hard. In the first weeks of the semester, when I had enough money for food, I would wake up early to revise notes before class; lectures always made sense to me, and I was sharp in seminars. But as my food money dwindled, I went hungry and could feel my attention span shrinking. I would not spend my time learning, but thinking of where to get my next meal. Instead of visiting the library, I would sleep. I would stay in my room rather than go out with other students — and I struggled in some of my courses.

At long last, universities and research institutions are starting to pay attention to bullying, harassment and mental health. Now, they need to recognize that far too many students in higher education are hungry or are spending their time worrying about where to get food. A survey at two universities in Nigeria found that 45% of students had gone hungry or cut down on their food consumption to save money — and even higher rates were found at a university in South Africa. Rich countries can also face this burden. The University of California estimates that one-quarter of its graduate students have experienced food insecurity, meaning that they have skipped meals or reduced portions to save money, or ran out of food before they could afford to buy more.

World Food Day is on 16 October, and I call on institutions of higher learning to address food insecurity on their campuses. I urge them to strategize around both long- and short-term solutions. It is humane as well as pragmatic to ensure that students can be fully present and actively learning in classrooms — which is impossible if they're too hungry.

As an agricultural researcher, I study beneficial soil microbes. My ultimate goal is to find sustainable ways to grow crops and prevent insect losses amid a changing climate. I have also established Oyeska Greens, an agriculture-focused start-up in Kwale, Kenya, that creates farming systems that produce more food using fewer resources than traditional farms. But I am increasingly aware that efficient food production is just one aspect, although perhaps the most straightforward, of creating a world with food security. For the benefit of their students — and to create a model for tackling important problems — educational institutions should take on the difficult task of making sure that nourishing food is available to the members of their communities.

Some institutions have taken the initiative. Several have food pantries or gardens on campus. The University of California, San Francisco, created an app to let students know when food is left over from catered events, and some 69% of its student population — all post-graduates — have signed up. The University of the Witwatersrand in Johannesburg, South Africa, set up its Food Sovereignty Centre and other outreach programmes to encourage student donations and to offer meals and food grown in a campus garden to matriculants in need.

> **BUT AS FOOD MONEY DWINDLED, I WENT HUNGRY AND COULD FEEL MY ATTENTION SPAN SHRINKING.**

And educational institutions must adopt a more comprehensive, long-term view. But how?

First, universities should collect hard data about hunger and food insecurity on campus. In 2018, the US Government Accountability Office found evidence that this was a growing problem, but that there was a dearth of data. Students already take surveys after completing courses and at key points in the academic year. Some of these should be co-opted, or new surveys should be commissioned, to address food security, so that educational institutions can assess how many students, postdocs and junior faculty members are worrying about hunger.

Even simple steps are useful, such as compiling lists of resources for students who face food insecurity, mental-health issues and other challenges. Cornell University in Ithaca, New York, and the University of Oregon in Eugene present this information in online letters to students. Accurate data could help to get effective messages to the most vulnerable.

Universities should also work to devise fresh ideas for tackling these issues. Students are the most affected, so institutions should engage with them to design solutions. I can imagine an innovation challenge that spans countries. Campuses could join together to share how they have solved or mitigated food insecurity and other challenges. Education leaders should record and monitor what makes campus programmes addressing this food insecurity sustainable through the years.

In the end, the hard truth is that combating hunger costs money. Universities should set aside funds to help students cope. At the same time, governments need to step up and create nutrition-assistance programmes for students, or at the very least ensure that students are eligible for existing ones.

The good news: change is happening. A coalition of more than 100 institutions across 29 countries asks students to take the lead and push administrators to fight hunger and food insecurity. That includes raising awareness, holding food drives and more.

Students can do much more than they or the societies they live in assume, and they should not be afraid to try. While I was a graduate student at Auburn University in Alabama, I founded a primary school in Kenya. It now serves more than 100 students from poor families. I know at first hand how difficult it is for children to learn when they are hungry. Because of my concern for these children, I made sure the school would provide them with meals — supplied in part by four greenhouses that grow food for the school and the community. When these students get to university and beyond, they will be all the more prepared to tackle the world's problems. ∎

**Esther Ngumbi** *is an assistant professor at the University of Illinois at Urbana-Champaign.*
*e-mail: est28@yahoo.com*

# SEVEN DAYS
*The news in brief*

## Huge Antarctic iceberg breaks loose

An iceberg bigger than Greater London has broken off Antarctica's third-largest ice shelf. The chunk of ice, which has an area of 1,636 square kilometres and weighs some 315 billion tonnes, came off the Amery Ice Shelf in East Antarctica on 26 September. Surprisingly, the berg, called D-28, broke off just west of an area that scientists had been watching more closely, which is dubbed the Loose Tooth because it looks ready to calve. The event is the latest in a cycle in which big icebergs break off the Amery shelf every six or seven decades. It is not thought to be linked to climate change, although other parts of Antarctica are experiencing rapid ice loss linked to warming. The calving of D-28 might affect how ice moves in this part of the Amery shelf, including the Loose Tooth region.

---

**POLICY**

## Trump vs science

US President Donald Trump's administration has driven government science into a full-blown crisis, according to a report released on 3 October by the Brennan Center for Justice at New York University. The authors, led by former US Attorney Preet Bharara and former Environmental Protection Agency administrator Christine Todd Whitman, say that political manipulation of science has reached unprecedented levels "with almost weekly violations of previously respected safeguards". They cite the Trump administration's decisions to disband federal advisory panels, suppress scientific reports and relocate or reassign government scientists, and the White House's efforts to circumvent Congress by allowing officials who have not been confirmed by the Senate to remain in leadership posts. The report calls for legislation that would establish stronger standards for scientific integrity and encourage the rapid appointment of qualified people to head crucial scientific posts.

**FACILITIES**

## Gravity agreement

Japan's Kamioka Gravitational-Wave Detector (KAGRA) near Hida has joined the international network of observatories that detect gravitational waves. KAGRA officials signed a memorandum of agreement on 4 October to pool data and publish joint results with the Laser Interferometer Gravitational-Wave Observatory (LIGO) in the United States and Virgo near Pisa in Italy; the combined power of the four interferometers (LIGO has two) will enhance the confidence and quality of each detection. KAGRA was completed this year and is due to start its first science run in December; with two 3-kilometre arms stretching through tunnels under a mountain, it is the world's first interferometer of its size to be built underground. It is also the first to run with cryogenic mirrors, cooled to 20 kelvin.

**SPACE**

## New moons

Astronomers have found 20 more moons orbiting Saturn, bringing the known total to 82 — the most in the Solar System. (Jupiter is second with 79.) Saturn's new-found moons, announced on 7 October, are each about 5 kilometres across. Seventeen of the 20 travel in a direction opposite to the planet's rotation, suggesting that they are fragments of a larger satellite that broke apart. One of these is the most distant known moon around Saturn. Two of the three other new-found moons travel in orbits that are similar to those of previously discovered moons. The third has an unusual, stretched-out trajectory. The discovery team, led by Scott Sheppard at the Carnegie Institution for Science in Washington DC, has launched a public naming contest. According to International Astronomical Union rules, the moons must be named after giants in Inuit, Norse or Gallic mythology.

**EVENTS**

## Indonesian fires

Rains in the past week have helped to extinguish intense wildfires in Indonesian peatlands that have burned for months. The Indonesian National Institute of Aeronautics and Space reported just

179 fire hotspots across the archipelago on 3 October; the previous month, there were 9,310 hotspots in the regions of Sumatra and Kalimantan alone. The Ministry of Environment and Forestry says that 328,000 hectares of forest and peatland burnt across the country between January and mid-September (**pictured**). "Almost 80,000 hectares of the burnings happened in peatlands," says Nazir Foead, the head of Indonesia's Peatland Restoration Agency, "and it created a thicker and prolonged haze." Government climatologists warn that the fires might come back, because the dry season lasts until the end of October.

## DNA testing

A US government plan to routinely collect DNA data from immigrants in federal custody is sparking concerns about privacy and discrimination. The Department of Homeland Security said on 2 October that it was developing regulations that would allow DNA profiles of all immigrant detainees to be stored in an FBI database created to help law-enforcement agencies solve violent crimes. The policy would affect the more than 40,000 people now being held in immigrant detention centres — as well as future detainees. Bioethicists say the efforts will wrongfully target an already vulnerable population. "To me, this is equivalent of a collection from an entire apartment complex where there's been a murder or where there might be crime," says Sara Katsanis, a bioethicist at Northwestern University in Chicago, Illinois, who opposes the government's plan.

## Nuclear-fusion plan

The United Kingdom has entered the race to build the world's first prototype commercial nuclear-fusion reactor, with a £200-million (US$270-million) investment announced by the government on 3 October. Over the next four years, the cash will fund the design of the Spherical Tokamak for Energy Production (STEP). In theory, the plant would produce hundreds of megawatts of net electricity from nuclear fusion — the process that powers the Sun — by the early 2040s, demonstrating that the technique is commercially viable. The cost of building the facility would stretch to billions of pounds. "It's ambitious and adventurous, but I think the fusion programme has to be," says Howard Wilson, director of the STEP programme at the UK Atomic Energy Authority, which runs the Culham Centre for Fusion Energy near Oxford and is leading the work. Several countries are planning prototype fusion reactors, but no facility has yet sustained a fusion reaction for long enough for it to generate more energy than it takes to run.

## Nobel prizes

A trio of researchers — William Kaelin, Peter Ratcliffe and Gregg Semenza — won the 2019 Nobel Prize in Physiology or Medicine on 7 October for describing how cells sense and respond to changing oxygen levels by switching genes on and off. Their discoveries, made in the 1990s, have been key in understanding human diseases such as cancer and anaemia. The Nobel physics prize, announced the following day, was awarded to cosmologist James Peebles and astronomers Michel Mayor and Didier Queloz for discoveries about the evolution of the Universe and Earth's place in it. In 1995, Mayor and his then-student Queloz made the first discovery of an extrasolar planet orbiting a Sun-like star. The field is now one of astronomy's hottest: more than 4,000 exoplanets have been detected. Peebles' theoretical work helped to establish the current 'standard model' of the evolution of the Universe. *Nature* went to press before the chemistry prize was announced. See pages 161 and 162 for more.

## TREND WATCH

Almost one in five vertebrate animals that live on land are traded on wildlife markets — a much greater proportion than previously thought. The findings come from one of the most comprehensive studies of the international wildlife trade yet, involving a survey of 31,745 species of mammal, bird, amphibian and reptile. The authors found the proportion of traded animals to be 40–60% higher than previous estimates had suggested, and predict that it could rise to more than one in four. To identify species that are currently traded, the researchers used databases maintained by the Convention on International Trade in Endangered Species of Wild Fauna and Flora, and the International Union for Conservation of Nature, whose Red List provides the conservation status of most species. They found that 5,579 of the species analysed — around 18% — are being bought and sold around the world. This includes more than 2,000 birds and nearly 1,500 mammals, many captured illegally from the wild, although the figure also includes legal trade. The authors say the findings could help to identify species at risk of extinction, so that policies can be put in place to protect them.

### VULNERABLE VERTEBRATES

Of 31,745 species — encompassing birds, mammals, amphibians and reptiles — around 18% have been reported as traded.
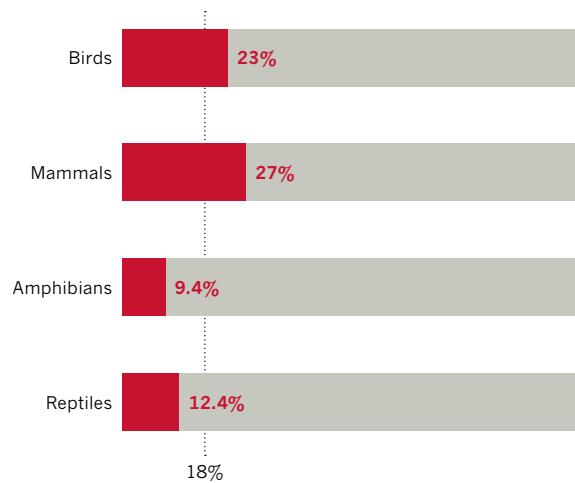
| | |
|---|---|
| Birds | 23% |
| Mammals | 27% |
| Amphibians | 9.4% |
| Reptiles | 12.4% |

18%

# NEWS IN FOCUS

Men extract a Guinea worm from the leg of an infected dog in Chad, where canine cases are driving the parasite's spread.

PUBLIC HEALTH

# Battle to wipe out Guinea worm stumbles

*World Health Organization delays target date for eradicating the parasite to 2030.*

**BY LESLIE ROBERTS**

A few years ago, it looked like humanity was about to wipe the debilitating parasite Guinea worm off the face of the Earth. But faced with evidence of previously unknown routes of transmission, the World Health Organization (WHO) has quietly pushed back the target date for eradication from 2020 to 2030.

"We are being realistic and down to earth," says Dieudonné Sankara, who heads the WHO's eradication effort.

So far, humanity has eradicated just one human pathogen: smallpox. The decision on Guinea worm (*Dracunculus medinensis*) is a major blow to the international partnership that has been fighting the parasite since the 1980s. Led by the Carter Center in Atlanta, Georgia, the partnership has reduced the

number of new infections from 3.5 million per year in 1986 to just 28 in 2018. The disease now lingers in a handful of Central African nations.

But a series of puzzling discoveries has made the 2020 target impossible to meet. The most urgent issue is the soaring and unexplained, rate of infection among dogs in Chad — which has helped to keep Guinea worm circulating. Then there are the first known cases among people in Angola, perplexing infections ▶

▶ in baboons in Ethiopia, and conflicts that have hampered eradication in Mali, Sudan and South Sudan. Some health experts wonder whether elimination of the parasite is possible.

"The question has been put on the table and has been on some of our minds," says Mark Eberhard, a retired parasitologist and member of a WHO advisory group whose job is to certify when Guinea worm is finally gone. He says that the dog cases suggest that eradication will be very difficult — if not impossible.

But Donald Hopkins, the tropical-medicine specialist who has led the Guinea-worm effort from its outset, is unwavering. "I am confident we will be able to wrap it up," says Hopkins, a special adviser at the Carter Center.

Guinea worm afflicts the poorest of the poor. There is no drug to treat it, and no vaccine to prevent it. People contract the parasite by drinking water that contains microscopic water fleas that carry Guinea-worm larvae. A year or so later, a worm 60 to 90 centimetres long erupts through the skin on the leg or foot. Its painful exit from the body can take weeks.

To relieve the burning sensation, many people wade into the nearest body of water. When an adult worm enters the water, it releases larvae, and the cycle starts anew.

For decades, parasitologists thought that this was the only route of transmission, and that the worm infected only people. Researchers devised a plan to eradicate the disease by teaching people to filter their drinking water and to stay out of ponds if a worm is emerging. Larvicide use complemented these measures.

The World Health Assembly endorsed the strategy in 1986. Experts thought that they could wipe out Guinea-worm disease because the parasite was not known to circulate in animals, which could help it to survive.

Those assumptions began to falter in 2010, when the disease popped up in people living along the Chari River in Chad after a ten-year absence. The cases were sporadic and dispersed, rather than clustered around contaminated water sources. Stranger still, eradication-programme staff spotted stringy worms hanging from the legs of domestic dogs. Genetic analysis confirmed that these parasites were *D. medinensis*, which had evaded surveillance in Chad for about a decade.

These developments suggested the existence of a route of transmission related to the fishing industry along the Chari River. But after eight years, researchers still haven't pinned it down. "What are we missing?" says Eberhard.

*"We are being realistic and down to earth."*

The number of new Guinea-worm infections in people has remained relatively constant in Chad, at about a dozen per year since 2010. Yet the number of new cases in dogs has climbed from hundreds in the early 2010s to more than 1,500 this year. "In Chad, it is clear that dogs are driving transmission," Eberhard says. "If we control it in dogs, human cases might go away."

Infected dogs have also been reported in Ethiopia and Mali, but the cases number in the tens and twenties, not the thousands seen in Chad. Researchers aren't sure why Chad has been hit so hard. "It is important that we understand more about the epidemiology of the disease — learn the really key source of infections in dogs," says Sarah Cleaveland, a veterinary surgeon and epidemiologist at the University of Glasgow, UK. She leads a WHO working group that is developing criteria to verify when animals are free of Guinea worm.

The discovery in 2013 of infected baboons in a small forested area in southern Ethiopia also has researchers scratching their heads. So far, scientists have found 15 baboons with Guinea-worm disease. A key question, Cleaveland says, is whether baboons, like dogs, can sustain transmission independently.

Then there is the emergence of Guinea-worm disease in Angola. In April 2018, an 8-year-old girl was diagnosed, followed by a second person and a dog this year.

"How long it has been there and where it came from is anyone's guess," Hopkins says. The parasite might have been lurking in Angola, or it could have hitched a ride into the country in a person or a dog. Scientists are looking for clues by sequencing DNA from Guinea-worm samples taken in Angola. The Carter Center is setting up surveillance in the country, and the WHO is working with the government of Namibia to scour its border with Angola for signs of the disease.

The WHO's new 2030 eradication target is intended to allow time not only to stop the transmission of Guinea worm, but to verify that the disease has gone. Doing so requires three or more years without an infection in a person or animal.

David Molyneux, a parasitologist at the Liverpool School of Tropical Medicine, UK, and a member of the WHO commission that will certify eradication, wonders how scientists will ever be sure that the worm has been vanquished. "Our job is to work out how you might certify a country the size of Chad free of dracunculiasis in humans and dogs. Can we ever envisage that level of surveillance?" he says.

He is pushing for a plan B in case wiping out Guinea worm proves impossible — and says that the world should celebrate what the eradication effort has already accomplished. "It has stopped millions of people from becoming disabled," he says.

But Hopkins is steadfast. "The daunting thing about eradication is there is no wiggle room," he says. "Zero is zero." ∎

---

DATA PRIVACY

# Facebook research hits a snag

*Sharing user data with external social scientists proves technically difficult.*

BY ELIZABETH GIBNEY

A pioneering research initiative designed to allow independent scientists to access Facebook data has run up against a major snag over privacy.

The project's goal was to enable academic researchers to study how social media is influencing democracies — and to establish a model of collaboration that would allow scientists to take advantage of tech companies' rich troves of data. But the funders backing the initiative are considering ending their support for the project because privacy issues have prevented Facebook from providing scientists with all the data that they were promised — and it's not clear when these might be made available.

Academic scientists are keen to get their hands on data from tech giants to conduct independent analyses, as concerns about misinformation on social-media sites plague political processes worldwide. The US-based research initiative — called the Social Media and Democracy Research Grants programme, launched in cooperation with Facebook last July — funded 12 projects that were designed to investigate topics such as the spread of fake news and how social media was used during elections in Italy, Chile and Germany.

But problems with the data quickly emerged: Facebook has been able to share some information with researchers, but providing them with more sensitive and detailed data without compromising user privacy proved technically more difficult than project organizers expected.

Last month, the eight charitable funders supporting the project gave Facebook until 30 September to provide the full data set; otherwise, they would wind up the programme.

Since the funders' statement, Facebook has released a further data set, but not the full range. Now that the deadline has passed, the Hewlett Foundation, one of the charities, says the group is assessing the next steps and determining which research proposals can be accomplished.

Other partners that have spent a year working with Facebook on data-sharing solutions say they are continuing their efforts to build a computing infrastructure that allows the company to share its data with researchers, irrespective of the funders' decisions. The partners will continue to release data sets in the coming weeks, and Facebook has more than 30 people working on the project, says Gary King, a social scientist at Harvard University in Cambridge, Massachusetts, and co-founder of Social Science One. This non-profit foundation was set up to act as a 'data broker' between Facebook and the researchers on this project and future initiatives. "To learn about societies, we must go to where the data are," says King. He says that the model his team is implementing is the only one plausible for collaborations with technology giants.

A spokesperson for Facebook told *Nature*: "This is one of the largest sets of links ever to be created for academic research on this topic. We are working hard to deliver on additional demographic fields while safeguarding individual people's privacy."

## DATA SHORTCOMINGS

At issue is the amount and type of information that Facebook has been able to give external researchers. Data sets released so far, for example, include 32 million links, or URLs, each of which has been shared publicly by at least 100 users. These include some valuable information, such as ratings of a page's trustworthiness from third-party fact-checking sites. But the company had promised researchers around one billion links, including those largely shared privately, where fake news tends to circulate, says Simon Hegelich, a political data scientist at the Technical University of Munich in Germany. His team is studying misinformation during Germany's 2017 election. "My impression is that, at least for our project, the data that Facebook is offering is more or less useless," he adds.

Sharing data with researchers without compromising user privacy required new infrastructure. Social Science One and Facebook built a secure portal that connects to Facebook's servers and uses a mathematical technique known as differential privacy, which adds noise to the results of analyses that prevents users from becoming personally identifiable. Social-media data, although less sensitive than medical information, bring extra privacy challenges because they are connected to a person's real-world behaviour, so even if they are anonymized, it is relatively easy to identify individuals, says Jake Metcalf, a technology ethicist at the think tank Data & Society in New York City who is on the team conducting ethical reviews for proposals to the scheme. "It's a very challenging model to achieve," he says. ∎



Ethiopia's highland waterfalls, where the Blue Nile begins.

WATER

# Nations clash over giant Nile dam

*Egypt says the Grand Ethiopian Renaissance Dam will cause water shortages — but Ethiopia stands firm.*

**BY ANTOANETA ROUSSI**

Environmental scientists representing Egypt, Ethiopia and Sudan are at the heart of an increasingly bitter dispute over Africa's largest hydroelectric dam, which Ethiopia is building on the Nile.

The countries' researchers met in Sudan's capital, Khartoum, ahead of a conference of water ministers on 4–5 October. The dam's environmental impacts, especially on water supplies in Egypt, topped the agenda. But the ministers' meeting ended without resolution and Egypt is now calling for the United States to become involved. Ethiopia opposes this.

Egypt is concerned that Ethiopia is moving too fast to complete the Grand Ethiopian Renaissance Dam, and that its timetable will create water and food scarcity and put millions of Egypt's farmers out of work. Ninety per cent of Egypt's fresh water comes from the Nile, which runs south to north from Ethiopia's highlands, the main source of the tributary called the Blue Nile.

Ethiopia counters that the project, which is 60% complete, is essential for its electricity needs and is a matter of national sovereignty — not something Egypt can interfere with.

According to the World Bank, 66% of Ethiopia's population is without electricity, the third highest proportion in the world. At its peak, the dam is expected to produce 6.45 gigawatts of electricity.

Ethiopia's government also says that its plan will enable countries to its north to cope more effectively with the effects of climate change. At present, unpredictable dry and wet weather in the Nile Basin — caused in part by climate change — is contributing to intermittent floods and water shortages. Ethiopia's plan will even out Nile water flow, making such events less likely, says Seleshi Bekele, Ethiopia's minister of water, irrigation and energy.

The three countries involved have established an independent expert panel, the National Independent Scientific Research Group, to help find a way forward.

## STARTING SCHEDULE

When the dam will start operating depends on how quickly its main reservoir can be filled from Nile water, and this is central to the dispute. The reservoir provides the store of water that is used to drive turbines and generate electricity. Ethiopia wants the reservoir to be filled over 5 years, with 35 billion cubic ▶

▶ metres of water being released to countries downstream each year while the dam is being filled. Egypt says that its water supplies will be reduced during this period. It is calling for the reservoir to be filled more slowly, over 7 years, and wants more water to be released — 40 billion cubic metres per year.

Egypt and Ethiopia do not have a formal water-sharing agreement. Under the 1959 Nile Waters Agreement between Egypt and Sudan, Egypt takes 55.5 billion cubic metres of water from the Nile each year, and Sudan takes 18.5 billion. The agreement was reached shortly before Egypt began constructing its own megadam, the Aswan High Dam (see 'A river runs through it').

Ethiopia, however, was not part of this agreement and therefore does not recognize it. Ethiopian foreign-ministry spokesperson Nebiyat Getachew said at a press conference on 20 September that any proposal that did not respect "Ethiopia's sovereignty and its right to use the Nile dam" wouldn't be accepted.

"Ethiopia expects discussions and progress on our talks without the imposition of any one of the countries," says Bekele. "The issues are solvable technically and we can place the right framework on long-term operation, based on science and best practices."

Water-resources researcher Kevin Wheeler of the University of Oxford, UK, says that in a year with average rainfall, Egypt should experience little or no extra water scarcity if the reservoir is filled over 5–7 years, with at least 35 billion cubic metres of water released downstream.

**A RIVER RUNS THROUGH IT**

Ethiopia, Egypt and Sudan are trying to resolve a dispute over Ethiopia's project to build Africa's largest hydropower dam.
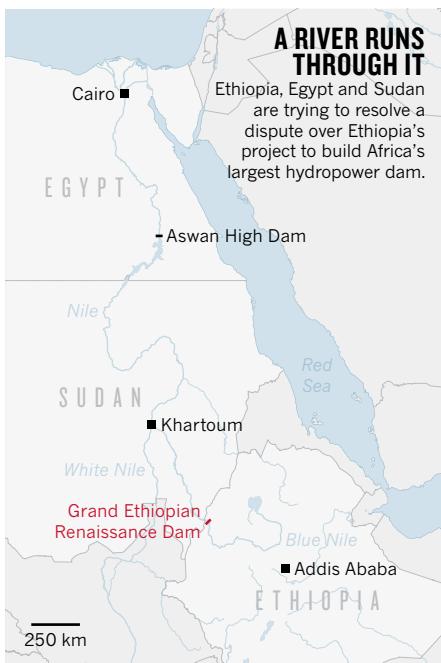
But Egypt is right to be concerned about extra water scarcity in dry years and those with low rainfall, adds Wheeler, who co-wrote a 2016 paper on ways to fill the dam (K. G. Wheeler *et al. Water Int.* **41**, 611–634; 2016).

Harry Verhoeven, a Nile Basin researcher based in Qatar, says that ultimately there is little Egypt can do, and policymakers in Cairo will have to adjust to having less Nile water during the dam's filling period. "Reduced water flows over several years mean tough choices, not only of who gets the water but what crops you grow and whether domestic food supply or export markets are prioritized," he says.

Verhoeven says that Egypt could take the dispute to the International Court of Justice in The Hague, the Netherlands, but that would require both sides to agree to such arbitration. Even if they did agree, he predicts, the court would be unlikely to find in Egypt's favour. "Ethiopia has a right to develop the water resources in its territory," he says.

Egypt's ministry of water and irrigation did not respond to *Nature*'s repeated requests for comment. But in a statement issued earlier this month, the ministry said that it considered "it important for the Ethiopian side to engage in serious technical negotiations", and find an agreement that would be in "the common interests of the three countries".

Although neither side has been willing to budge so far, the countries are likely to find a compromise, says Ismail Serageldin, a former vice-president of the World Bank who predicted in 1995 that twenty-first-century wars would be fought over water. "Ethiopia wants as short a period as possible, Egypt wants as long a period as possible, they will negotiate and meet somewhere in the middle — I think it's good that people are talking."

"There's still time for wars," adds Serageldin, who later became a science adviser to Egypt's prime minister. "But who knows, we may turn out to be wise; wiser than I thought possible at the time that I said that." ∎

# Stratospheric data aid climate forecasts

*Including the upper atmosphere in weather models helped understanding of rare Antarctic event.*

**BY DYANI LEWIS**

For the past month, a rare atmospheric phenomenon has been brewing above Antarctica, raising temperatures in the upper atmosphere by 40 degrees and threatening to reverse the direction of a powerful jet stream for only the second time since records began.

At the first signs of this event, known as sudden stratospheric warming, Eun-Pa Lim, a climate scientist at the Australian Bureau of Meteorology in Melbourne, plugged the rising temperatures into a model she had designed that forecasts short-term climate over the Southern Hemisphere (E.-P. Lim *et al. J. Geophys. Res. Atmos.* **123**, 12002–12016; 2018). The model predicted that the warming above Antarctica will drive hot, dry winds across eastern Australia over the next three months.

The forecast has excited meteorologists because it shows how far the field has come in understanding the stratosphere — the second major layer of Earth's atmosphere — and its effects on weather.

For decades, meteorologists thought weather was mostly driven by what was happening in the troposphere, the layer between the stratosphere and Earth's surface. Then, in 2001, daily stratospheric weather maps revealed how the two regions interact (M. P. Baldwin and T. J. Dunkerton *Science* **294**, 581–584; 2001). Now these interactions are being included in models such as the one designed by Lim to forecast short-term climate — conditions occurring between a 7–10-day weather forecast and the following 3 months — around the world. For instance, meteorologists can now predict how conditions in the stratosphere will affect a climatic phenomenon that drives heavy rainfall in the United States in winter.

"We have a much better understanding of how the stratosphere affects the weather at the surface," says Adam Scaife, head of long-range forecasting at the Met Office Hadley Centre for Climate Science and Services in Exeter, UK.

Improved accuracy and confidence in such forecasts makes a big difference to government agencies preparing for heatwaves or fires, as well as to farmers, such as those in drought-affected eastern Australia, when planning irrigation or herd-mustering schedules, says Lim.

Sudden stratospheric warming events are common in the Northern Hemisphere, occurring every second year, on average, but they are rare in the Southern Hemisphere. The first such event recorded in the south, in 2002, took scientists by surprise.

Even if they had known it was coming,

models back then couldn't have predicted how the abrupt warming in the stratosphere might affect the weather, says Harry Hendon, head of climate processes at the Australian Bureau of Meteorology.

Climate models have improved significantly over the past 15 years, partly driven by faster, cheaper computers. They're also much better at combining sources of observational data, such as satellite measurements of stratospheric temperature and atmospheric humidity.

Such advances helped meteorologists to forecast the start of the current stratospheric warming about a week in advance. The events typically start towards the end of winter, when mountains or the contrast between warm ocean temperatures and cold land masses generate continental-scale atmospheric disturbances known as Rossby waves. If these are large enough, they can reach into the stratosphere and break like a wave over a beach, compressing and warming the air in the stratosphere above the pole. This pressure can force the strong stratospheric winds encircling the pole — the polar-night jet stream — to abruptly slow and reverse, changing from being westerly winds to flowing in an easterly direction, says Scaife.

A complete reversal has not yet occurred in the current event, but wind speeds have already plummeted. Scientists at the Bureau of Meteorology don't know exactly what sparked this year's event, but they predict that it will be stronger than in 2002 — and so have a greater effect on the weather.

Lim's model, which teases out how stratospheric conditions bleed down into the troposphere, has helped to predict how this might play out. Apart from bringing warmer weather to eastern Australia, the event will drive colder, wetter conditions to western Tasmania, New Zealand's South Island and the southern tip of South America.

The warming so far has also sent an influx of ozone-rich air to counter the thinning of ozone over Antarctica that usually occurs in spring.

Meteorologists are now waiting to see whether the forecast holds. Hendon hopes that, if it does, the bureau will incorporate Lim's model into its standard operations, to provide short-term climate predictions every spring. ∎

---

MEDICINE PRIZE

# Biologists who decoded oxygen sensing win Nobel

*Laureates' discovery underpins understanding of diseases such as anaemia and cancer.*

**BY HEIDI LEDFORD & EWEN CALLAWAY**

A trio of researchers has won the 2019 Nobel Prize in Physiology or Medicine for describing how cells sense and respond to changing oxygen levels by switching genes on and off — a discovery that has been key in understanding human diseases such as cancer and anaemia.

The three scientists are cancer researcher William Kaelin at the Dana-Farber Cancer Institute in Boston, Massachusetts; physician–scientist Peter Ratcliffe at the University of Oxford, UK, and the Francis Crick Institute in London; and geneticist Gregg Semenza at Johns Hopkins University in Baltimore, Maryland. The team also won the Albert Lasker Basic Medical Research Award in 2016.

Their work has helped researchers to understand how the body adapts to low oxygen levels by, for example, cranking out red blood cells and growing new blood vessels.

"This is a fundamental discovery that they've contributed to," says Celeste Simon, a cancer biologist at the University of Pennsylvania in Philadelphia. "All organisms need oxygen, so it's really important."

"The field really coalesced around this discovery, which was dependent on each one of their findings," says Randall Johnson, a physiologist at the University of Cambridge, UK, and the Karolinska Institute in Stockholm, and a member of the Nobel Assembly. "This really was a three-legged stool."

**OXYGEN DEPRIVATION**
The body's tissues can be deprived of oxygen during exercise or when blood flow is interrupted, such as during a stroke. Cells' ability to sense oxygen is also crucial for the developing fetus and placenta, as well as for tumour growth, because the mass of rapidly growing cells can deplete oxygen in a tumour's interior.

In work conducted in the 1990s, the scientists discovered the molecular processes that cells go through to respond to oxygen levels in the body. They found that central to this is a mechanism involving proteins called hypoxia-inducible factor (HIF) and VHL.

Semenza and Ratcliffe studied the regulation of a hormone called erythropoietin (EPO), which is crucial for stimulating the production of red blood cells in response to low oxygen levels. Semenza and his team identified a pair of genes that encode the two proteins that form the protein complex HIF, which turns on certain genes and boosts EPO production when oxygen is low.

Meanwhile, Kaelin showed that a gene called *VHL* also seemed to be involved in how cells respond to oxygen. Kaelin was studying a genetic syndrome called von Hippel-Lindau's disease; families with the disease carry mutations in *VHL*, and the condition raises the risk of certain cancers.

**Nobel prizewinners Peter Ratcliffe (left), William Kaelin (centre) and Gregg Semenza (right).**

▶ Ratcliffe and his team later found that the protein expressed by *VHL* interacts with a component of HIF, turning off responses to low-oxygen conditions by marking the HIF component for destruction once oxygen levels rise.

And in 2001, teams led by Kaelin and Ratcliffe elucidated more details about this process. They discovered that, when oxygen is present, a chemical modification to the VHL protein called prolyl hydroxylation allows VHL to bind HIF, which leads to the latter's breakdown. But this modification is blocked when cells are oxygen-starved, kick-starting the activity of HIF.

As a result, cells can react to low oxygen levels by simply blocking the breakdown of HIF, notes Mark Dewhirst, a cancer biologist at Duke University in Durham, North Carolina. "The cell can respond in minutes."

## DRUG DEVELOPMENT

The work has led researchers to develop drugs that target oxygen-sensing processes, including those in cancer. Drugs, called prolyl hydroxylase inhibitors, that prevent VHL from binding to HIF and causing its degradation are also being investigated as treatments for anaemia and renal failure. Chinese regulators approved the first of these drugs in 2018.

"You could argue that some aspect of this is going to be germane to all diseases you can think of," says Simon.

Colleagues hailed the trio as role models for other scientists. "They are extremely humble people," says Dewhirst. "All three of them hold scientific rigour and reproducibility to the absolute highest standard," adds Simon.

Kaelin, in particular, has taken his field to task for pursuing possible cancer treatments that aren't backed up by strong evidence. "The most dangerous result in science is the one you were hoping for, because you declare victory and get lazy," he told scientists at a 2018 talk at the US National Institutes of Health in Bethesda, Maryland. ∎

---



Didier Queloz (left), James Peebles (centre) and Michel Mayor.

**PHYSICS PRIZE**

# Planet pioneers win physics Nobel

*Exoplanet astronomers share award with cosmologist whose theories describe Universe's evolution.*

**BY ELIZABETH GIBNEY & DAVIDE CASTELVECCHI**

Cosmologist James Peebles and astronomers Michel Mayor and Didier Queloz have won the 2019 Nobel Prize in Physics for discoveries about the evolution of the Universe and Earth's place in it.

In 1995, Mayor, at the University of Geneva, Switzerland, and his then-student Queloz made the first discovery of a planet orbiting a Sun-like star (M. Mayor and D. Queloz *Nature* **378**, 355–359; 1995). Their work launched a field that has become one of astronomy's hottest. They detected the exoplanet through its tiny gravitational pull on its star, 51 Pegasi, a technique that is now used to study some of the more than 4,000 exoplanets now known to exist.

Peebles, who is at Princeton University in New Jersey, developed a theoretical framework that underpins modern understanding of the Universe's history (P. J. E. Peebles and J. T. Yu *Astrophys. J.* **162**, 815; 1970). In particular, he helped to lay the theoretical foundations for the cosmic microwave background (CMB), the 'afterglow' of the Big Bang, and to establish the current 'standard model' of the Universe's evolution. In this model, the mysterious substance known as dark matter plays a central part in assembling large-scale structures of the cosmos, such as galaxies and clusters of galaxies.

Mayor and Queloz share one half of the prize, worth 9 million Swedish kronor (US$910,000), and Peebles will receive the other half.

Mayor and Queloz's discovery "started modern exoplanet science", says Guillem Anglada-Escudé, an astronomer at the Institute for Space Sciences-CSIC in Barcelona, Spain.

Researchers had discovered exoplanets orbiting spinning cores of dead stars known as pulsars, but not around stars similar to our own, which could host habitable planets. The pair's discovery came as a surprise. The planet they detected, called 51 Pegasi b, is a gas giant, a type that astronomers had expected would orbit the outer reaches of a solar system. But it was orbiting much closer to its star than Mercury is to the Sun — an early sign that other planetary systems might not be like our own.

The finding was remarkable for being almost completely unambiguous and quickly confirmed, says Anglada-Escudé.

## PROBING FIRST LIGHT

Meanwhile, Peebles' theories have allowed cosmologists to understand much more about the CMB and the Universe's beginnings.

"Were it not for the theoretical discoveries of James Peebles, the wonderful high-precision measurements of this radiation over the last 20 years would have told us almost nothing," said Mats Larsson, a molecular physicist at Stockholm University and chair of the 2019 Nobel physics committee, when he revealed the prize.

Peebles developed a model of the Universe's evolution known as the 'cold dark matter' theory, which describes how cosmological structures formed as the Universe expanded and cooled from its hot, dense beginnings. Together with the later addition of ideas about dark energy, this has become the standard framework of modern cosmology.

Although the precise nature of dark matter has yet to be understood, several high-precision surveys of the Universe have lent support to this theory; these include studies of the CMB and the mapping of galaxies across large swathes of the sky. "This is such a long-deserved recognition," says François Bouchet, an astronomer at the Institute of Astrophysics in Paris.

It is unusual for exoplanets and cosmology to be paired up in the same prize, but both lines of work "give a fresh perspective of the place humans have in the cosmos", says Bouchet. ∎

# DEEP TROUBLE FOR DEEP LEARNING

BY DOUGLAS HEAVEN

## ARTIFICIAL-INTELLIGENCE RESEARCHERS ARE TRYING TO FIX THE FLAWS OF NEURAL NETWORKS.

A self-driving car approaches a stop sign, but instead of slowing down, it accelerates into the busy intersection. An accident report later reveals that four small rectangles had been stuck to the face of the sign. These fooled the car's onboard artificial intelligence (AI) into misreading the word 'stop' as 'speed limit 45'.

Such an event hasn't actually happened, but the potential for sabotaging AI is very real. Researchers have already demonstrated how to fool an AI system into misreading a stop sign, by carefully positioning stickers on it[1]. They have deceived facial-recognition systems by sticking a printed pattern on glasses or hats. And they have tricked speech-recognition systems into hearing phantom phrases by inserting patterns of white noise in the audio.

These are just some examples of how easy it is to break the leading pattern-recognition technology in AI, known as deep neural networks (DNNs). These have proved incredibly successful at correctly classifying all kinds of input, including images, speech and data on consumer preferences. They are part of daily life, running

# FOOLING THE AI

Deep neural networks (DNNs) are brilliant at image recognition — but they can be easily hacked.

These stickers made an artificial-intelligence system read this stop sign as 'speed limit 45'.

**Stop** → **Speed limit 45**
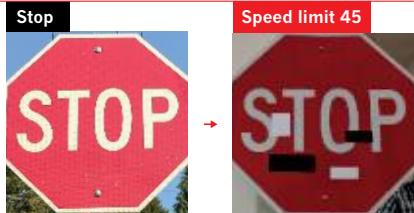
Scientists have evolved images that look like abstract patterns — but which DNNs see as familiar objects.

**King penguin** | **Starfish**

Adding carefully crafted noise to a picture can create a new image that people would see as identical, but which a DNN sees as utterly different.

**Panda** + **Gibbon**
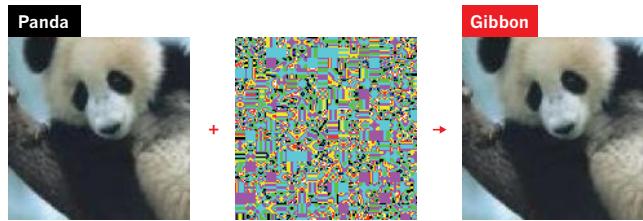
In this way, any starting image can be tweaked so a DNN misclassifies it as any target image a researcher chooses.

**Sloth** + **Target image: race car** → **Race car**

Rotating objects in an image confuses DNNs, probably because they are too different from the types of image used to train the network.

**Stop** → **Dumb-bell** → **Racket**

Even natural images can fool a DNN, because it might focus on the picture's colour, texture or background rather than picking out the salient features a human would recognize.

**Manhole cover** | **Pretzel**

everything from automated telephone systems to user recommendations on the streaming service Netflix. Yet making alterations to inputs — in the form of tiny changes that are typically imperceptible to humans — can flummox the best neural networks around.

These problems are more concerning than idiosyncratic quirks in a not-quite-perfect technology, says Dan Hendrycks, a PhD student in computer science at the University of California, Berkeley. Like many scientists, he has come to see them as the most striking illustration that DNNs are fundamentally brittle: brilliant at what they do until, taken into unfamiliar territory, they break in unpredictable ways (see 'Fooling the AI').

That could lead to substantial problems. Deep-learning systems are increasingly moving out of the lab into the real world, from piloting self-driving cars to mapping crime and diagnosing disease. But pixels maliciously added to medical scans could fool a DNN into wrongly detecting cancer, one study reported this year[2]. Another suggested that a hacker could use these weaknesses to hijack an online AI-based system so that it runs the invader's own algorithms[3].

In their efforts to work out what's going wrong, researchers have discovered a lot about why DNNs fail. "There are no fixes for the fundamental brittleness of deep neural networks," argues François Chollet, an AI engineer at Google in Mountain View, California. To move beyond the flaws, he and others say, researchers need to augment pattern-matching DNNs with extra abilities: for instance, making AIs that can explore the world for themselves, write their own code and retain memories. These kinds of system will, some experts think, form the story of the coming decade in AI research.

## REALITY CHECK

In 2011, Google revealed a system that could recognize cats in YouTube videos, and soon after came a wave of DNN-based classification systems. "Everybody was saying, 'Wow, this is amazing, computers are finally able to understand the world,'" says Jeff Clune at the University of Wyoming in Laramie, who is also a senior research manager at Uber AI Labs in San Francisco, California.

But AI researchers knew that DNNs do not actually understand the world. Loosely modelled on the architecture of the brain, they are software structures made up of large numbers of digital neurons arranged in many layers. Each neuron is connected to others in layers above and below it.

The idea is that features of the raw input coming into the bottom layers — such as pixels in an image — trigger some of those neurons, which then pass on a signal to neurons in the layer above according to simple mathematical rules. Training a DNN network involves exposing it to a massive collection of examples, each time tweaking the way in which the neurons are connected so that, eventually, the top layer gives the desired answer — such as always interpreting a picture of a lion as a lion, even if the DNN hasn't seen that picture before.

A first big reality check came in 2013, when Google researcher Christian Szegedy and his colleagues posted a preprint called 'Intriguing properties of neural networks'[4]. The team showed that it was possible to take an image — of a lion, for example — that a DNN could identify and, by altering a few pixels, convince the machine that it was looking at something different, such as a library. The team called the doctored images 'adversarial examples'.

A year later, Clune and his then-PhD student Anh Nguyen, together with Jason Yosinski at Cornell University in Ithaca, New York, showed that it was possible to make DNNs see things that were not there, such as a penguin in a pattern of wavy lines[5]. "Anybody who has played with machine learning knows these systems make stupid mistakes once in a while," says Yoshua Bengio at the University of Montreal in Canada, who is a pioneer of deep learning. "What was a surprise was the type of mistake," he says. "That was pretty striking. It's a type of mistake we would not have imagined would happen."

New types of mistake have come thick and fast. Last year, Nguyen, who is now at Auburn University in Alabama, showed that simply rotating objects in an image was sufficient to throw off some of the best image classifiers around[6]. This year, Hendrycks and his colleagues reported

that even unadulterated, natural images can still trick state-of-the-art classifiers into making unpredictable gaffes, such as identifying a mushroom as a pretzel or a dragonfly as a manhole cover[7].

The issue goes beyond object recognition: any AI that uses DNNs to classify inputs — such as speech — can be fooled. AIs that play games can be sabotaged: in 2017, computer scientist Sandy Huang, a PhD student at the University of California, Berkeley, and her colleagues focused on DNNs that had been trained to beat Atari video games through a process called reinforcement learning[8]. In this approach, an AI is given a goal and, in response to a range of inputs, learns through trial and error what to do to reach that goal. It is the technology behind superhuman game-playing AIs such as AlphaZero and the poker bot Pluribus. Even so, Huang's team was able to make their AIs lose games by adding one or two random pixels to the screen.

Earlier this year, AI PhD student Adam Gleave at the University of California, Berkeley, and his colleagues demonstrated that it is possible to introduce an agent to an AI's environment that acts out an 'adversarial policy' designed to confuse the AI's responses[9]. For example, an AI footballer trained to kick a ball past an AI goalkeeper in a simulated environment loses its ability to score when the goalkeeper starts to behave in unexpected ways, such as collapsing on the ground.

> ## "THERE ARE SO MANY DIFFERENT WAYS THAT YOU CAN ATTACK A SYSTEM."

Knowing where a DNN's weak spots are could even let a hacker take over a powerful AI. One example of that came last year, when a team from Google showed that it was possible to use adversarial examples not only to force a DNN to make specific mistakes, but also to reprogram it entirely — effectively repurposing an AI trained on one task to do another[3].

Many neural networks, such as those that learn to understand language, can, in principle, be used to encode any other computer program. "In theory, you can turn a chatbot into whatever programme you want," says Clune. "This is where the mind starts to boggle." He imagines a situation in the near future in which hackers could hijack neural nets in the cloud to run their own spambot-dodging algorithms.

For computer scientist Dawn Song at the University of California, Berkeley, DNNs are like sitting ducks. "There are so many different ways that you can attack a system," she says. "And defence is very, very difficult."

## WITH GREAT POWER COMES GREAT FRAGILITY

DNNs are powerful because their many layers mean they can pick up on patterns in many different features of an input when attempting to classify it. An AI trained to recognize aircraft might find that features such as patches of colour, texture or background are just as strong predictors as the things that we would consider salient, such as wings. But this also means that a very small change in the input can tip it over into what the AI considers an apparently different state.

One answer is simply to throw more data at the AI; in particular, to repeatedly expose the AI to problematic cases and correct its errors. In this form of 'adversarial training', as one network learns to identify objects, a second tries to change the first network's inputs so that it makes mistakes. In this way, adversarial examples become part of a DNN's training data.

Hendrycks and his colleagues have suggested quantifying a DNN's robustness against making errors by testing how it performs against a large range of adversarial examples. However, training a network to withstand one kind of attack could weaken it against others, they say. And researchers led by Pushmeet Kohli at Google DeepMind in London are trying to inoculate DNNs against making mistakes. Many adversarial attacks work by making tiny tweaks to the component parts of an input — such as subtly altering the colour of pixels in an image — until this tips a DNN over into a misclassification. Kohli's team has suggested

that a robust DNN should not change its output as a result of small changes in its input, and that this property might be mathematically incorporated into the network, constraining how it learns.

For the moment, however, no one has a fix on the overall problem of brittle AIs. The root of the issue, says Bengio, is that DNNs don't have a good model of how to pick out what matters. When an AI sees a doctored image of a lion as a library, a person still sees a lion because they have a mental model of the animal that rests on a set of high-level features — ears, a tail, a mane and so on — that lets them abstract away from low-level arbitrary or incidental details. "We know from prior experience which features are the salient ones," says Bengio. "And that comes from a deep understanding of the structure of the world."

One attempt to address this is to combine DNNs with symbolic AI, which was the dominant paradigm in AI before machine learning. With symbolic AI, machines reasoned using hard-coded rules about how the world worked, such as that it contains discrete objects and that they are related to one another in various ways. Some researchers, such as psychologist Gary Marcus at New York University, say hybrid AI models are the way forward. "Deep learning is so useful in the short term that people have lost sight of the long term," says Marcus, who is a long-time critic of the current deep-learning approach. In May, he co-founded a start-up called Robust AI in Palo Alto, California, which aims to mix deep learning with rule-based AI techniques to develop robots that can operate safely alongside people. Exactly what the company is working on remains under wraps.

Even if rules can be embedded into DNNs, they are still only as good as the data they learn from. Bengio says that AI agents need to learn in richer environments that they can explore. For example, most computer-vision systems fail to recognize that a can of beer is cylindrical because they were trained on data sets of 2D images. That is why Nguyen and colleagues found it so easy to fool DNNs by presenting familiar objects from different perspectives. Learning in a 3D environment — real or simulated — will help.

But the way AIs do their learning also needs to change. "Learning about causality needs to be done by agents that do things in the world, that can experiment and explore," says Bengio. Another deep-learning pioneer, Jürgen Schmidhuber at the Dalle Molle Institute for Artificial Intelligence Research in Manno, Switzerland, thinks along similar lines. Pattern recognition is extremely powerful, he says — good enough to have made companies such as Alibaba, Tencent, Amazon, Facebook and Google the most valuable in the world. "But there's a much bigger wave coming," he says. "And this will be about machines that manipulate the world and create their own data through their own actions."

In a sense, AIs that use reinforcement learning to beat computer games are doing this already in artificial environments: by trial and error, they manipulate pixels on screen in allowed ways until they reach a goal. But real environments are much richer than the simulated or curated data sets on which most DNNs train today.

## ROBOTS THAT IMPROVISE

In a laboratory at the University of California, Berkeley, a robot arm rummages through clutter. It picks up a red bowl and uses it to nudge a blue oven glove a couple of centimetres to the right. It drops the bowl and picks up an empty plastic spray bottle. Then it explores the heft and shape of a paperback book. Over several days of non-stop sifting, the robot starts to get a feel for these alien objects and what it can do with them.

The robot arm is using deep learning to teach itself to use tools. Given a tray of objects, it picks up and looks at each in turn, seeing what happens when it moves them around and knocks one object into another.

When researchers give the robot a goal — for instance, presenting it with an image of a nearly empty tray and specifying that the robot arrange objects to match that state — it improvises, and can work with objects it has not seen before, such as using a sponge to wipe objects off a table. It also figured out that clearing up using a plastic water bottle to knock objects out of the way is quicker than picking up those objects directly. "Compared to other machine-learning techniques, the generality of what it can accomplish continues to impress me," says

Chelsea Finn, who worked at the Berkeley lab and is now continuing that research at Stanford University in California.

This kind of learning gives an AI a much richer understanding of objects and the world in general, says Finn. If you had seen a water bottle or a sponge only in photographs, you might be able to recognize them in other images. But you would not really understand what they were or what they could be used for. "Your understanding of the world would be much shallower than if you could actually interact with them," she says.

But this learning is a slow process. In a simulated environment, an AI can rattle through examples at lightning speed. In 2017, AlphaZero, the latest version of DeepMind's self-taught game-playing software, was trained to become a superhuman player of Go, then chess and then shogi (a form of Japanese chess) in just over a day. In that time, it played more than 20 million training games of each event.

AI robots can't learn this quickly. Almost all major results in deep learning have relied heavily on large amounts of data, says Jeff Mahler, co-founder of Ambidextrous, an AI and robotics company in Berkeley, California. "Collecting tens of millions of data points would cost years of continuous execution time on a single robot." What's more, the data might not be reliable, because the calibration of sensors can change over time and hardware can degrade.

Because of this, most robotics work that involves deep learning still uses simulated environments to speed up the training. "What you can learn depends on how good the simulators are," says David Kent, a PhD student in robotics at the Georgia Institute of Technology in Atlanta. Simulators are improving all the time, and researchers are getting better at transferring lessons learnt in virtual worlds over to the real. Such simulations are still no match for real-world complexities, however.

Finn argues that learning using robots is ultimately easier to scale up than learning with artificial data. Her tool-using robot took a few days to learn a relatively simple task, but it did not require heavy monitoring. "You just run the robot and just kind of check in with it every once in a while," she says. She imagines one day having lots of robots out in the world left to their own devices, learning around the clock. This should be possible — after all, this is how people gain an understanding of the world. "A baby doesn't learn by downloading data from Facebook," says Schmidhuber.

### LEARNING FROM LESS DATA

A baby can also recognize new examples from just a few data points: even if they have never seen a giraffe before, they can still learn to spot one after seeing it once or twice. Part of the reason this works so quickly is because the baby has seen many other living things, if not giraffes, so is already familiar with their salient features.

A catch-all term for granting these kinds of abilities to AIs is transfer learning: the idea being to transfer the knowledge gained from previous rounds of training to another task. One way to do this is to reuse all or part of a pre-trained network as the starting point when training for a new task. For example, reusing parts of a DNN that has already been trained to identify one type of animal — such as those layers that recognize basic body shape — could give a new network the edge when learning to identify a giraffe.

An extreme form of transfer learning aims to train a new network by showing it just a handful of examples, and sometimes only one. Known as one-shot or few-shot learning, this relies heavily on pre-trained DNNs. Imagine you want to build a facial-recognition system that identifies people in a criminal database. A quick way is to use a DNN that has already seen millions of faces (not necessarily those in the database) so that it has a good idea of salient features, such as the shapes of noses and jaws. Now, when the network looks at just one instance of a new face, it can extract a useful feature set from that image. It can then compare how similar that feature set is to those of single images in the criminal database, and find the closest match.

Having a pre-trained memory of this kind can help AIs to recognize new examples without needing to see lots of patterns, which could speed up learning with robots. But such DNNs might still be at a loss when confronted with anything too far from their experience.

It's still not clear how much these networks can generalize.

Even the most successful AI systems such as DeepMind's AlphaZero have an extremely narrow sphere of expertise. AlphaZero's algorithm can be trained to play both Go and chess, but not both at once. Retraining a model's connections and responses so that it can win at chess resets any previous experience it had of Go. "If you think about it from the perspective of a human, this is kind of ridiculous," says Finn. People don't forget what they've learnt so easily.

### LEARNING HOW TO LEARN

AlphaZero's success at playing games wasn't just down to effective reinforcement learning, but also to an algorithm that helped it (using a variant of a technique called Monte Carlo tree search) to narrow down its choices from the possible next steps[10]. In other words, the AI was guided in how best to learn from its environment. Chollet thinks that an important next step in AI will be to give DNNs the ability to write their own such algorithms, rather than using code provided by humans.

Supplementing basic pattern-matching with reasoning abilities would make AIs better at dealing with inputs beyond their comfort zone, he argues. Computer scientists have for years studied program synthesis, in which a computer generates code automatically. Combining that field with deep learning could lead to systems with DNNs that are much closer to the abstract mental models that humans use, Chollet thinks.

> ## "A BABY DOESN'T LEARN BY DOWNLOADING DATA FROM FACEBOOK."

In robotics, for instance, computer scientist Kristen Grauman at Facebook AI Research in Menlo Park, California, and the University of Texas at Austin is teaching robots how best to explore new environments for themselves. This can involve picking in which directions to look when presented with new scenes, for instance, and which way to manipulate an object to best understand its shape or purpose. The idea is to get the AI to predict which new viewpoint or angle will give it the most useful new data to learn from.

Researchers in the field say they are making progress in fixing deep learning's flaws, but acknowledge that they're still groping for new techniques to make the process less brittle. There is not much theory behind deep learning, says Song. "If something doesn't work, it's difficult to figure out why," she says. "The whole field is still very empirical. You just have to try things."

For the moment, although scientists recognize the brittleness of DNNs and their reliance on large amounts of data, most say that the technique is here to stay. The realization this decade that neural networks — allied with enormous computing resources — can be trained to recognize patterns so well remains a revelation. "No one really has any idea how to better it," says Clune. ∎

**Douglas Heaven** *is a freelance writer based in London.*

1. Eykholt, K. *et al. IEEE/CVF Conf. Comp. Vision Pattern Recog.* **2018**, 1625–1634 (2018).
2. Finlayson, S. G. *et al. Science* **363**, 1287–1289 (2019).
3. Elsayed, G. F., Goodfellow, I. & Sohl-Dickstein, J. Preprint at https://arxiv.org/abs/1806.11146 (2018).
4. Szegedy, C. *et al.* Preprint at https://arxiv.org/abs/1312.6199v1 (2013).
5. Nguyen, A., Yosinski, J. & Clune, J. *IEEE Conf. Comp. Vision Pattern Recog.* **2015**, 427–436 (2015).
6. Alcorn, M. A. *et al. IEEE Conf. Comp. Vision Pattern Recog.* **2019**, 4845–4854 (2019).
7. Hendrycks, D., Zhao, K., Basart, S., Steinhardt, J. & Song, D. Preprint at https://arxiv.org/abs/1907.07174 (2019).
8. Huang, S., Papernot, N., Goodfellow, I., Duan, Y. & Abbeel, P. Preprint at https://arxiv.org/abs/1702.02284 (2017).
9. Gleave, A. *et al*. Preprint at https://arxiv.org/abs/1905.10615 (2019).
10. Silver, D. *et al. Science* **362**, 1140–1144 (2018).

# COMMENT

ILLUSTRATION BY SEÑOR SALME

# How science has shifted our sense of identity

Biological advances have repeatedly changed who we think we are, writes **Nathaniel Comfort**, in the third essay of a series on how the past 150 years have shaped science.

In the iconic frontispiece to Thomas Henry Huxley's *Evidence as to Man's Place in Nature* (1863), primate skeletons march across the page and, presumably, into the future: "Gibbon, Orang, Chimpanzee, Gorilla, Man." Fresh evidence from anatomy and palaeontology had made humans' place on the scala naturae scientifically irrefutable.

We were unequivocally with the animals — albeit at the head of the line.

Nicolaus Copernicus had displaced

**150 YEARS OF NATURE**
Anniversary collection
go.nature.com/nature150

us from the centre of the Universe; now Charles Darwin had displaced us from the centre of the living world. Regardless of how one took this demotion (Huxley wasn't troubled; Darwin was), there was no doubting Huxley's larger message: science alone can answer what he called the 'question of questions': "Man's place in nature and his ▶

relations to the Universe of things."

Huxley's question had a prominent place in the early issues of *Nature* magazine. Witty and provocative, 'Darwin's bulldog' was among the most in-demand essayists of the day. Norman Lockyer, the magazine's founding editor, scored a coup when he persuaded his friend to become a regular contributor. And Huxley knew a soapbox when he saw one. He hopped up and used *Nature*'s pages to make his case for Darwinism and the public utility of science.

It was in the seventh issue — 16 December 1869 — that Huxley advanced a scheme for what he called 'practical Darwinism' and we call eugenics. Convinced that continued dominance of the British Empire would depend on the "energetic enterprising" English character, he mused about selecting for a can-do attitude among Britons[1]. Acknowledging that the law, not to mention ethics, might get in the way, he nevertheless wrote: "it may be possible, indirectly, to influence the character and prosperity of our descendants." Francis Galton — Darwin's cousin and an outer planet of Huxley's solar system — was already writing about similar ideas and would come to be known as the father of eugenics. When this magazine appeared, then, the idea of 'improving' human heredity was on many people's minds — not least as a potent tool of empire.

Huxley's sunny view — of infinite human progress and triumph, brought about by the inexorable march of science — epitomizes a problem with so-called Enlightenment values. The precept that society should be based on reason, facts and universal truths has been a guiding theme of modern times. Which in many ways is a splendid thing (lately I've seen enough governance without facts for one lifetime). Yet Occam's razor is double edged. Enlightenment values have accommodated screechingly discordant beliefs, such as that all men are created equal, that aristocrats should be decapitated and that people can be traded as chattel.

I want to suggest that many of the worst chapters of this history result from scientism: the ideology that science is the only valid way to understand the world and solve social problems. Where science has often expanded and liberated our sense of self, scientism has constrained it.

Across the arc of the past 150 years, we can see both science and scientism shaping human identity in many ways. Developmental psychology zeroed in on the intellect, leading to the transformation of IQ (intelligence quotient) from an educational tool into a weapon of social control. Immunology redefined the 'self' in terms of 'non-self'. Information theory provided fresh metaphors that recast identity as residing in a text or a wiring diagram. More recently, cell and molecular studies have relaxed the borders of the self. Reproductive technology, genetic engineering and synthetic biology
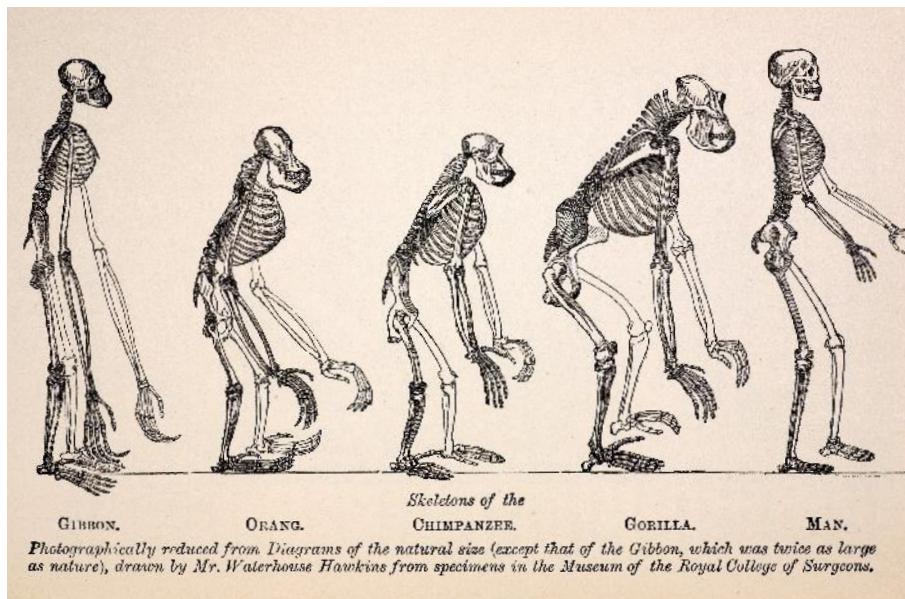


Frontispiece to Thomas Henry Huxley's *Evidence as to Man's Place in Nature* (1863).

have made human nature more malleable, epigenetics and microbiology complicate notions of individuality and autonomy, and biotechnology and information technology suggest a world where the self is distributed, dispersed, atomized.

Individual identities, rooted in biology, have perhaps never played a larger part in social life, even as their bounds and parameters grow ever fuzzier.

## DESIGNS ON INTELLIGENCE

"Methods of scientific precision must be introduced into all educational work, to carry everywhere good sense and light," wrote the French psychologist Alfred Binet in 1914 (ref. 2). A decade earlier, Binet and Théodore Simon developed a series of tests for French schoolchildren to measure what they called 'mental age'. If a child's mental age was less than her chronological age, she could receive extra help to catch up. The German psychologist William Stern took the ratio of mental to chronological age, giving what he called the IQ and, theoretically, making it comparable across groups. Meanwhile, Charles Spearman, a British statistician and eugenicist of the Galton school, found a correlation between a child's performance on different tests. To explain the correlations, he theorized an innate, fixed, underlying quality he called 'g', for 'general intelligence'. Then the American psychologist Henry Goddard, with the eugenicist Charles Davenport whispering in his ear, claimed that low IQ was a simple Mendelian trait. Thus, step by scientistic step, IQ was converted from a measure of a given

*"Information theory provided fresh metaphors that recast identity as residing in a text or a wiring diagram."*

child's past performance to a predictor of any child's future performance.

IQ became a measure not of what you do, but of who you are — a score for one's inherent worth as a person. In the Progressive era, eugenicists became obsessed with low intelligence, believing it to be the root of crime, poverty, promiscuity and disease. By the time Adolf Hitler expanded eugenics to cover entire ethnic and cultural groups, tens of thousands of people worldwide had already been yanked from the gene pool, sterilized, institutionalized, or both.

## NOT ME

Immunologists took another approach, They located identity in the body, defining it in relational rather than absolute terms: self and non-self. Tissue-graft rejection, allergies and autoimmune reactions could be understood not as a war but as an identity crisis. This was pretty philosophical territory. Indeed, the historian Warwick Anderson has suggested that[3] in immunology, biological and social thought have been "mixing promiscuously in a common tropical setting, under the palm trees".

The immunological Plato was the Australian immunologist Frank MacFarlane Burnet. Burnet's fashioning of immunology as the science of the self was a direct response to his reading of the philosopher Alfred North Whitehead. Tit for tat, social theorists from Jacques Derrida to Bruno Latour and Donna Haraway have leaned on immunological imagery and concepts in theorizing the self in society. The point is that scientific and social thought are deeply entangled, resonant, co-constructed. You can't fully understand one without the other.

Later, Burnet was drawn to new metaphors taken from cybernetics and information theory. "It is in the spirit of the times," he wrote in 1954 (ref. 4), to believe there would soon

be "a 'communications theory' of the living organism." Indeed there was. In the same period, molecular biologists also became enamoured of information metaphors. After the 1953 solution of the DNA double helix, as the problem of the genetic code took shape, molecular biologists found analogies with information, text and communication irresistible, borrowing words such as 'transcription', 'translation', 'messengers', 'transfers' and 'signalling'. The genome 'spells' in an 'alphabet' of four letters, and is almost invariably discussed as a text, whether it is a book, manual or parts list. Not coincidentally, these fields grew up alongside computer science and the computing industry.

The postwar self became a cipher to be decoded. DNA sequences could be digitized. Its messages could, at least in theory, be intercepted, decoded and programmed. Soon it became hard not to think of human nature in terms of information. By the 1960s, DNA was becoming known as the 'secret of life'.

## MANY SELVES
In the late 1960s and 1970s, critics (including a number of scientists) grew concerned that the new biology could alter what it means to be human. The ethical and social issues raised were "far too important to be left solely in the hands of the scientific and medical communities", wrote James Watson (of DNA fame and later infamy) in 1971.

In 1978, Patrick Steptoe and Robert Edwards succeeded with human *in vitro* fertilization, leading to the birth of Louise Brown, the first 'test-tube baby'. By 1996, human cloning seemed to be around the corner, with the cloning of a sheep that Ian Wilmut and his team named Dolly.

Cloning and genetic engineering have prompted much soul-searching but little soul-finding. There has long been something both terrible and fascinating about the idea of a human-made, perhaps not-quite-person. Would a cloned individual have the same rights as the naturally born? Would a baby conceived or engineered to be a tissue donor be somehow dehumanized? Do we have a right to alter the genes of the unborn? Or, as provocateurs have argued, do we have an obligation to do so? The recent development of potent gene-editing tools such as CRISPR has only made widening participation in such decision-making more urgent.

Arguments, both pro and con, around engineering humans often lean on an overly deterministic understanding of genetic identity. Scientism can cut both ways. A deep reductionism located human nature inside the cell nucleus. In 1902, the English physician Archibald Garrod had written[5] of genetically based "chemical individuality". In the 1990s, as the first tsunamis of genomic sequence data began to wash up on the shores of basic science, it became obvious that human genetic variation was much

more extensive than we had realized. Garrod has become a totem of the genome age.

By the end of the century, visionaries had begun to tout the coming of 'personalized medicine' based on your genome. No more 'one size fits all', went the slogan. Instead, diagnostics and therapy would be tailored to you — that is, to your DNA. After the Human Genome Project, the cost of DNA sequencing nosedived, making 'getting your genome done' part of mass culture.

Today, tech-forward colleges offer genome profiles to all incoming first-years. Hip companies purport to use your genome to compose personalized wine lists, nutritional supplements, skin cream, smoothies or lip balm. The sequence has become the self. As it says on the DNA testing kit from sequencing company 23andMe, "Welcome to you."

## BOUNDARIES BLUR
But you are not all you — not by a long shot. The DNA-as-blueprint model is outdated, almost quaint. For starters, all of the cells in a body do not have the same chromosomes. Cisgender women are mosaics: the random inactivation of one X chromosome in each cell means that half a woman's cells express her mother's X and half express her father's. Mothers are also chimaeras, thanks to the exchange of cells with a fetus through the placenta.

Chimaerism can cross the species boundary, too. Human–chimpanzee embryos have been made in the laboratory, and researchers are hard at work trying to grow immune-tolerant human organs in pigs. Genes, proteins and microorganisms stream continuously among almost any life forms living cheek by jowl. John Lennon was right: "I am he as you are he as you are me and we are all together."

*"Autoimmune reactions could be understood not as a war but as an identity crisis."*

Even in strictly scientific terms, 'you' are more than the contents of your chromosomes. The human body contains at least as many non-human cells (mostly bacteria, archaea and fungi) as human ones[6]. Tens of thousands of microbial species crowd and jostle over and through the body, with profound effects on digestion, complexion, disease resistance, vision and mood. Without them, you don't feel like you; in fact, you aren't really you. The biological self has been reframed as a cluster of communities, all in communication with each other.

These, too, cavort promiscuously beneath the palms. Scientists found that they could use a person's microbiome to identify their sexual partner 86% of the time[7]. The communities of greatest similarity in cohabiting couples, they found, are on the feet. The thigh microbiome, by contrast, is more closely correlated with your biological sex

than with the identity of your partner.

A body part, a cesspool, a subway car, a classroom — any place with a characteristic community — can be understood as having a genetic identity. In such a community, genetic information passes within and between individual organisms, through sex, predation, infection and horizontal gene transfer. In the past year, studies have shown that the communities of symbiotic microbes in deep-sea mussels become genetically isolated over time, like species. In fungi, genes called *Spok* (spore-killer) ebb and flow and recombine across species by 'meiotic drive', a kind of genomic fast-forward button that permits heritable genetic change to occur fast enough to respond to a rapidly changing environment. The genome, as the geneticist Barbara McClintock said long ago, is a sensitive organ of the cell.

Epigenetics dissolves the boundaries of the self even further. Messages coded in the DNA can be modified in many ways — by mixing and matching DNA modules, by capping or hiding bits so that they can't be read, or by changing the message after it's been read, its meaning altered in translation. DNA was once taught as a sacred text handed faithfully down the generations. Now, increasing evidence points to the nuclear genome as more of a grab bag of suggestions, tourist phrases, syllables and gibberish that you use and modify as needed. The genome now seems less like the seat of the self and more of a toolkit for fashioning the self. So who is doing the fashioning?

## DISTRIBUTED SELF
Brain implants, human–machine interfaces and other neurotechnical devices extend the self into the domain of the 'universe of things'. Elon Musk's company Neuralink in San Francisco, California, seeks to make the seamless mind–machine interface — that sci-fi trope — a (virtual) reality. Natural intelligence and artificial intelligence already meet; it's not far-fetched for them to somehow, someday, meld.

Can the self become not merely extended but distributed? The writer and former *Nature* editor Philip Ball let researchers sample his skin cells, turn them back into stem cells (with the potential to become any organ) and then culture them into a 'mini-brain', neural tissue in a dish that developed electrical firing patterns typical of regions of the brain. Other sci-fi staples, such as growing whole brains in Petri dishes or culturing human organs in farm animals, remain a long way off, but active efforts to achieve them are under way.

## SELF CONTROL
Yet there is a fruit fly in the ointment. Most of these Age-of-Reason notions of identity, and the dominant sci-fi scenarios of post-human futures, have been developed by university-educated men who were not

A macaque undergoing a liver transplant from a pig in China in 2013.

disabled, and who hailed from the middle and upper classes of wealthy nations of the global north. Their ideas reflect not only the findings but also the values of those who have for too long commanded the science system: positivist, reductionist and focused on dominating nature. Those who control the means of sequence production get to write the story.

That has begun to change. Although there is far to go, greater attention to equity, inclusion and diversity has already profoundly shaped thinking about disease, health and what it means to be human. It matters that Henrietta Lacks, whose tumour cells are used in labs all over the world, cultured and distributed without her consent, was a poor African American woman. Her story has stimulated countless conversations about inequities and biases in biomedicine, and changed practices at the United States' largest biomedical funder, the National Institutes of Health.

Considering genomic genealogy from an African American perspective, the sociologist Alondra Nelson has revealed complex, emotionally charged efforts to recover family histories lost to the Middle Passage. In the Native American community, creation of a genetic Native identity was a co-production of Western science and Indigenous culture, as the historian Kim TallBear has shown.

DNA-based conceptions of ethnicity are far from unproblematic. But the impulse to make the technologies of the self more accessible, more democratic — more about self-determination and less about social control — is, at its basis, liberatory.

Nowhere is this clearer than for people living with disabilities and using assistive technologies. They might gain or regain modes of perception, might be able to communicate and express themselves in new ways, and gain new relationships to the universe of things.

The artist Lisa Park plays with these ideas. She uses biofeedback and sensor technologies derived from neuroscience to create what she calls audiovisual representations of the self. A tree of light blooms and dazzles as viewers hold hands; pools of water resonate harmonically in response to Park's electroencephalogram waves; an 'orchestra' of cyborg musicians wearing heart and brain sensors make eerily beautiful music by reacting and interacting in different ways as Park, the conductor, instructs them to remove blindfolds, gaze at one another, wink, laugh, touch or kiss. Yet even this artistic, subjective and interactive

*"Evidence points to the nuclear genome as more of a grab bag of suggestions."*

sense of self is tied to an identity bounded by biology.

Since the Enlightenment, we have tended to define human identity and worth in terms of the values of science itself, as if it alone could tell us who we are. That is an odd and blinkered notion. In the face of colonialism, slavery, opioid epidemics, environmental degradation and climate change, the idea that Western science and technology are the only reliable sources of self-knowledge is no longer tenable. This isn't to lay all human misery at science's feet — far from it. The problem is scientism. Defining the self only in biological terms tends to obscure other forms of identity, such as one's labour or social role. Maybe the answer to Huxley's 'question of questions' isn't a number, after all. ∎

**Nathaniel Comfort** *is Professor of the History of Medicine at Johns Hopkins University, Baltimore, Maryland.*
*e-mail: nccomfort@gmail.com*

1. H. [Huxley, T. H.] *Nature* **1**, 183–184 (1869).
2. Binet, A. & Simon, T. *Mentally Defective Children* (Arnold, 1914).
3. Anderson, W. *Isis* **105**, 606–616 (2014).
4. Burnet, M. *Sci. Am.* **191**, 74–78 (1954).
5. Garrod, A. E. *Lancet* **160**, 1616–1620 (1902).
6. Sender, R., Fuchs, S. & Milo, R. *Cell* **164**, 337–340 (2016).
7. Ross, A. A., Doxey, A. C. & Neufeld, J. D. *mSystems* **2**, e00043-17 (2017).

*Every doctor in private practice was asked:*
—family physicians, surgeons, specialists…
doctors in every branch of medicine—
*"What cigarette do you smoke?"*

**A US magazine advertisement from around 1950.**

PUBLIC HEALTH

# The tobacco wars 2.0

**Felicity Lawrence** praises a history of the habit that kills eight million people a year.

The history of the tobacco industry, and its shameful campaign to delay regulation while millions died because of its products, might seem fully explored. Yet in her chronicle *The Cigarette*, historian Sarah Milov manages to bring fresh insight into how the industry's power hooked government treasuries, the advertising business and scientists for hire, to trump public health for so long. Tobacco killed an estimated 100 million people in the twentieth century. Without radical action, it is projected to kill around one billion in the twenty-first.

Many others have entered this reeking territory. They include journalist Richard Kluger, whose book *Ashes to Ashes* (1996) exposed the tobacco denial machine through hundreds of interviews with apologists and critics. In *The Cigarette Century* (2007), medical historian Allan Brandt interrogated cultural, scientific, legal and political evidence to explain how the industry created a global pandemic. *The Golden Holocaust* (2011) by science historian Robert Proctor



**The Cigarette: A Political History**
SARAH MILOV
*Harvard University Press (2019)*

mined millions of industry documents disclosed during litigation to produce an impassioned indictment of 'big tobacco', its plots and collaborators. Collectively, these catalogues of conspiracy go a long way towards explaining the persistence of smoking, decades after its potentially fatal impact was firmly established in the early 1960s.

What Milov adds is a nuanced account of the interplay between corporate machinations and government support for the industry from the 1930s until very recently. US state bureaucracies in tobacco-growing areas, and organizations such as the Farm Bureau that represented tobacco farmers in those states, are put forward as

co-conspirators. Her focus is the United States, but the arguments apply to the global industry. And the parallels with, say, the spread of junk food long linked to obesity are all too clear — with companies using the same strategies and even the same lobby groups.

**CORPORATE CONSPIRACY**
Much has been written about the tobacco industry deliberately obscuring the effects of smoking, not least by Naomi Oreskes and Eric Conway in the 2010 *Merchants of Doubt*. But during the First World War, the US federal government turned tobacco merchant itself. Classifying the industry as essential, it authorized the inclusion of rolling papers and tobacco in troops' rations. When the Second World War presented another industrial crisis, the government stepped in again. Britain had stopped importing US cigarettes, to conserve foreign currency for its war effort. So the US government bought the volumes equivalent

to the UK export market, to protect its own farmers.

The government had been bailing out tobacco farmers since the 1930s. The federal price-support system for tobacco began with the 1933 Agricultural Adjustment Act, part of president Franklin D. Roosevelt's New Deal to combat the Great Depression. In 1964, US surgeon-general Luther Terry released the report 'Smoking and Health', concluding that smoking caused premature death from lung cancer, emphysema, bronchitis and coronary heart disease. Yet it was not until 2004 that federal price support was terminated, even though nearly half a million US citizens continued to die from tobacco-related deaths yearly. (Government payments to tobacco farmers continued until 2014, to soften the blow.)

## MANIPULATING THE MARKET

During the cold war, Milov recounts, mass consumption of cigarettes was promoted by a burgeoning advertising industry. Smoking came to symbolize the triumph of consumer capitalism's abundance over the dreary shortages of Soviet socialism. It was in this context that the Tobacco Associates was established in 1947. A marketing board to promote the sale of US surplus cigarettes overseas, it was a private organization mandated by government to collect a levy from industry to fund its efforts.

This intertwined private and public policy effort — "associationalism", to use the jargon of political economists — had a key role in spreading the global epidemic of smoking-related diseases. By 1955 in the United States, more than half of all men and nearly one-quarter of all women smoked. Finding new smokers in other countries was seen as key to continued growth. It still is.

The US Marshall Plan to rebuild a devastated Europe after the Second World War had included loans to buy US tobacco as well as food. From 1954 onwards, the plan evolved into the Public Law 480 programme of aid to allies, increasingly in southeast Asia, Latin America and the Middle East. More often known as the Food for Peace programme, it gave tobacco preferential terms, alongside food. The result was, as intended, the establishment of permanent export markets for US commodities and the building of US geopolitical hegemony.

The effective counterpunch began in the late 1960s. When anti-smoking campaigners wanted to take on the nexus of industry, producer and state interests, they found two main routes to success. First, activists worked out how to harness the civil- and consumer-rights movements of the 1960s and 1970s to shift public perceptions of smoking and make it socially unacceptable. Young lawyer John Banzhaf, who founded the campaign group Action on Smoking and Health (ASH) in 1967, found ways to sue



In debates over vaping, tropes generated by tobacco interests are re-emerging.

the industry. Faced with a federal legislature that sided with manufacturing and farming lobbies, campaigners took their fight to local government, where the corporate lobbying machine was less established. They worked with city administrations and specialist regulators, achieving a ban on US broadcast advertisements for cigarettes in 1971, and restrictions on smoking on aircraft in 1973 through the Civil Aeronautics Board.

## GRASS-ROOTS ACTIVISM

Working alongside them were grass-roots activists, on whom Milov has fascinating detail. Clara Gouin, a Maryland woman with a child allergic to cigarette smoke, founded GASP — the Group Against Smokers' Pollution — in her living room in 1971. With others, she created the concept of the non-smoker, whose rights in public spaces were just as important as the smoker's.

The second front in the fight was the push to prove that smoking was economically damaging — and not just to governments picking up health bills. Smoking harmed productivity. In 1976, another woman, Donna Shimp, brought the first case against an employer over smokers in her workplace making her ill. She went on to make the

business case for banning tobacco in the workplace.

As ever, the callousness of tobacco's defenders continues to shock. It is impossible not to be outraged by the Tobacco Industry Research Committee, a formal conspiracy between cigarette manufacturers, agreed in a hotel room in December 1953. The group spent more than US$300 million between 1954 and 1997 on manufacturing doubt about the science on smoking and health.

The World Health Organization estimates that more than eight million people still die each year from smoking. This is happening even as the same old tropes return in debates over vaping, following deaths among people using e-cigarettes. Weeds, as Milov puts it, are hard to kill. Meanwhile, try substituting fossil fuels and climate change for tobacco and premature death in this history. You will find the same outrageous industry efforts to subvert science — and the same glimmers of hope for a counter-movement. ∎

**Felicity Lawrence** *is special correspondent for* The Guardian *in London and author of* Not on the Label *and* Eat Your Heart Out. *e-mail: felicity.lawrence@theguardian.com*

**Doris Lessing, photographed in 1990.**

FICTION

# Doris Lessing at 100: roving time and space

On the Nobel laureate's centenary, **Patrick French** explores her science–infused series Canopus in Argos.

In the 1920s, growing up on a poor farm in Southern Rhodesia (now Zimbabwe) Doris Lessing received impromptu outdoor lessons in space science from her mother, Emily. "Stones stood for Pluto, for Mars. I was Mercury and my brother Venus, running around my father, while she was the earth, moving slowly," she wrote.

Lessing wove space exploration, migration, climate change and social disintegration into novels that seem astonishingly prescient today. Leaving Africa for London in 1949, the fierce intelligence and impulsive curiosity of this driven autodidact led her into literary investigations of subjects as varied as genetics, nuclear warfare and post-colonialism. She wrote more than 60 books between 1950 and her death in 2013 at the age of 94. Six years before she died, she won the Nobel Prize in Literature — the first, and so far only, British woman to do so.

Her lifelong interest in science and societal upheaval is embodied in fascinating ways in Canopus in Argos, a series of five books published from 1979 to 1983. (She came up with the title a few weeks after seeing, and loving, George Lucas's film *Star*

*Wars*, in 1978. The inspiration might have been the 'crawl text' at the film's start.)

Lessing intended the first book, *Shikasta*, to break the bounds of her earlier work. She wanted to write open-ended space fiction as a study of social systems, taking in colonial dominance, sexuality and gender, evolution, eschatology and ideas about memory and power. She was not very interested in the mechanics of science fiction: a character might be "space-lifted" to another planet with little explanation. But in her futuristic anthropological analysis, much else of sci-fi culture is recognizable. She had been writing psychedelic, semi-realist fiction a decade before: *The Four-Gated City* (1969) ends in plague and the outbreak of the Third World War.

Today, as I research Lessing's authorized biography, I rarely find readers who appreciate both the space-based and the earthbound books. Admirers of the Canopus series, who tend to be younger and from a scientific background, usually have little interest in the rest of her work. And I have heard literary fans of *The Golden Notebook* (1962) declare in pained terms how much they resent her wilful shift to sci-fi.

*Shikasta* presents a revised version of Earth in the titular planet. Reports written by colonial servants of the galactic empire Canopus, historical texts, accounts and case studies create a diffuse narrative. There are echoes of Southern Rhodesia, where white settlers had seeded themselves at the end of the nineteenth century: Lessing described it as "a very nasty little police state". For instance, native people on Rohanda (a colony that reappears in the more sociological third book, *The Sirian Experiments*) are subjected to "an all-out booster, Top-Level Priority, Forced-Growth Plan", an explicit imperialist project. The second book, *The Marriages Between Zones Three, Four and Five,* set in 'zones' of civilization circling Shikasta, is an intense and explosive exploration of gender dynamics and stereotypical interactions between men and women.

In *The Making of the Representative for Planet 8*, Lessing examines human behaviour in the face of a brutal ice age. The inhabitants of Planet 8 must ultimately accept climate-based extinction, aided by a Canopean official, Doeg. This mythic apocalyptic parable was influenced by Anna Kavan's 1967 sci-fi novel *Ice*, as well as the death of British explorer Robert Falcon Scott in Antarctica in 1912, which fascinated Lessing. Towards the end, echoing current perceptions of planetary fragility, an inhabitant notes that "what we were seeing now with our new eyes was that all the planet had become a fine frail web or lattice, with the spaces held there between the patterns of the atoms".

The last book, *The Sentimental Agents in the Volyen Empire*, is a tonal shift: a farcical study of imperialism or, as Lessing asserted, "old-fashioned satire in space fiction terms".

## BITING BACK

Lessing wrote the five books at typically breakneck speed. Initially, they were greeted with bafflement. Novelist Anthony Burgess, author of the dystopian novel *A Clockwork Orange* (1962), complained of her "fanciful cosmic viewpoint". Although science-fiction doyenne Ursula K. Le Guin praised some character sketches in *Shikasta* as "immortal diamond", she found the whole at times "little more than a pulp-Galactic Empire with the Goodies fighting the Baddies".

Undeterred, Lessing worked her way through the series, declaring bloody-mindedly that "space fiction, with science fiction, makes up the most original branch of literature now". She had friends among sci-fi authors, including Brian Aldiss, and happily attended meetings of the International Conference on the Fantastic in the Arts. She championed the genre as influential in mainstream literature, whose pundits nevertheless "are much to blame for patronising or ignoring it". The critical readings became more analytical by 1982, when she published *The Making of the Representative for Planet 8*,

the series' most moving book.

Canopus in Argos also offered Lessing a way to address her own past, present and alternative futures. She had long presumed that a nuclear bomb was likely to fall on Europe, and that the planet faced annihilation. In 1957, she was present at the formation of the Campaign for Nuclear Disarmament. She believed in extraterrestrial intelligent life, and collected information on NASA plans for a 'man-in-a-can' hybrid spacecraft and rover; a 'complete defence shelter system' for nuclear attacks made by German company Thyssen; and schematics of space colonies and ventilation systems prepared for NASA's space scientists.

## SPACE FLIGHT

Her interest in space persisted. In the 1980s, she wrote the libretti for US composer Philip Glass's grand operatic adaptation of *The Making of the Representative for Planet 8* (she collaborated again with him on his 1997 opera *The Marriages Between Zones Three, Four, and Five*). In 1988, Glass arranged a visit to NASA's spaceflight centre in Houston, Texas, where she toured a model of the first US space station, Skylab, with John Frassanito, who had helped to design its interiors.

When a respected novelist veers off on a new path, critics will seek to find the intellectual rationale. They see it as a set of deliberate choices, and this interpretation can be stoked by the writer offering confident justifications in interviews, as Lessing did for Canopus in Argos. The biographer, by contrast, tends to search for proximate personal causes, tying the shift to moments of psychological importance for the writer.

And with Lessing, the biographical aspect is important. For instance, along with her interest in scientific fields from physics to neurology, she shared and influenced the counter-cultural mood prevailing among young people in the 1960s. By the late 1970s this became doom-laden, in response to environmental threats such as toxic waste. Youthful revolt over planetary destruction permeates *Shikasta* in particular.

Yet Lessing resisted classification. Her speculative space fiction was part of an unusual creative journey. Her next two novels turned from space back to Earth, and youth to age. Written under the pseudonym Jane Somers, they pivoted on the state of elderhood — which now, on our greying planet, has become another burgeoning field of study. ∎

**Patrick French** *is dean of the School of Arts and Sciences and professor for the public understanding of the humanities at Ahmedabad University in India. He is the authorized biographer of Doris Lessing. His books include* The World Is What It Is: The Authorized Biography of V.S. Naipaul. *e-mail: prbfrench@gmail.com*

# Books in brief
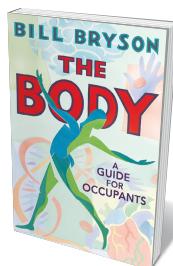
### The Body
*Bill Bryson* DOUBLEDAY *(2019)*
From skin to gut, the human body is a realm of wonder, and Bill Bryson's tome explores it to its thrumming depths. The book bristles with data such as our allotment of cells (37.2 trillion) or daily faeces production (200 grams), but the star turns are Bryson's wry forays into the histories of neuroscience, genetics, anatomy and immunology. Cue visceral gems such as diarist Samuel Pepys's gruesome bladder-stone surgery, and US physician Chevalier Quixote Jackson's retrieval of thousands of ingested items (including miniature binoculars and a poker chip) over his 75-year career.
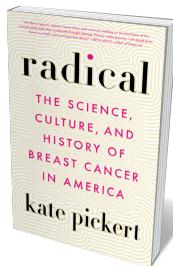
### Radical
*Kate Pickert* LITTLE, BROWN SPARK *(2019)*
Part of Kate Pickert's beat as a health-care journalist was breast cancer. In 2014, she became one of 300,000 US women diagnosed with the condition that year, and set out to recontextualize its convoluted history. She probes the brutal legacy of controversial mastectomy pioneer William Halsted, the discovery of cancer drug Taxol (paclitaxel) and debates over screening. She tours pharmaceuticals giant Genentech, interviews researchers such as Dennis Slamon and sits in on breast-reconstruction surgery. And she recounts her own medical journey with impressive aplomb. Balanced, cogent and eye-opening.
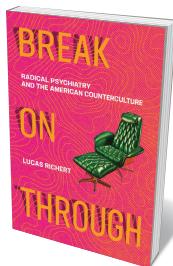
### Break on Through
*Lucas Richert* MIT PRESS *(2019)*
Sixty years ago, amid socio-economic stresses and cultural convulsions, US psychiatry went through a paradigm shift: radical approaches to therapy, newly approved pharmaceuticals and experimentation with hallucinogens proliferated. In this episodic narrative, historian of pharmacy Lucas Richert picks through the explosive developments alongside the multitude of figures involved, such as psychologist Abraham Maslow, anti-psychiatrist R. D. Laing, ex-patient and activist Judi Chamberlin and researcher Sanford M. Unger, who studied the use of LSD in psychotherapy.

### The Art of Innovation
*Ian Blatchford and Tilly Blyth* BANTAM *(2019)*
This fascinating compilation of 20 "brief yet rich" historical moments when art and science commingled draws on a BBC Radio 4 series by Ian Blatchford and Tilly Blyth. Director and principal curator at London's Science Museum, respectively, they gaze back over 250 years of crossover creativity. Here are landscape painter John Constable "skying" in the 1820s, painting cloudscapes and jotting down meteorological data; the mind-boggling motion photography of Eadweard Muybridge and Étienne-Jules Marey; and the mathematical models that inspired sculptor Barbara Hepworth.
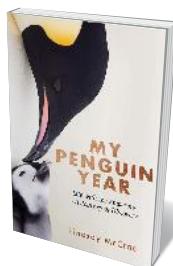
### My Penguin Year
*Lindsay McCrae* HODDER & STOUGHTON *(2019)*
In December 2016, Lindsay McCrae set out for Antarctica as director of photography for the BBC television series *Dynasties*, narrated by David Attenborough. Amid ice, whales, petrels, seals and vast shoals of fish, McCrae followed thousands of emperor penguins (*Aptenodytes forsteri*) for nearly a year. His remarkable memoir is rich in the technological and logistical challenges of filming in extreme conditions. But most gripping are his fine-tuned observations of these beautiful metre-high birds, which must survive and raise their young in temperatures as low as −60 °C. **Barbara Kiser**

# Correspondence

## AI behaviour: don't reinvent the wheel

The call of Iyad Rahwan and colleagues for a science of "machine behaviour" that empirically studies artificial intelligence (AI) "in the wild" (*Nature* **568**, 477–486; 2019) is an example of 'columbusing'. That is, what they claim to have discovered is, in fact, an existing field of study that has been producing vibrant, engaged research for decades. Cybernetics, the science of communications and automatic control systems in machines and living things, has been flourishing since the 1940s.

In our view, this prior art exposes serious ethical and scientific problems with the authors' proposal. Studying AI agents as if they are animate moves responsibility for the behaviour of machines away from their designers, thereby undermining efforts to establish professional ethics codes for AI practitioners.

The authors' idea that those who create machine-learning systems and study their behaviour cannot anticipate their "downstream societal effects" is false. Sociologists and anthropologists have long contributed to research on AI. For example, social scientists have described how AI can embed human intentions in material infrastructures (W. E. Bijker *et al.* (eds) *The Social Construction of Technological Systems*; 2012). Most would foresee AI agents' societal outcomes.

Columbusing fails to give due credit. It rides roughshod over long-fought struggles to centre science and technology's ethical implications for crucial issues such as inclusivity and diversity. All too often, those struggles have been fought by women and individuals of colour, who have laid much of the overlooked intellectual foundations of their disciplines.

**Emanuel Moss*** *Data & Society, New York City, New York, USA.*

*On behalf of 6 correspondents; see go.nature.com/2r5cjin. emanuel@datasociety.net

## 'Productivity' can be twisted: it's political

Oliver Hauser and colleagues' model of economics and game theory uses a technical parameter that they call 'productivity' (*Nature* **572**, 524–527; 2019). This introduces an ambiguity that has political implications because it does not align with the usual meaning of productivity when applied to income inequality.

In the model, individuals can each contribute some portion of their allocated resources to public goods that pay out to all participants. The twist is that the multiplier between donated resources and societal payout can vary from individual to individual. This multiplier is referred to as 'productivity', a term that, with respect to income inequality, conventionally implies individuals with large economic output. The multiplier in Hauser and colleagues' model refers instead to returns on the portion of invested resource — and only if they are donated back to create public goods.

Hauser *et al.* conclude that the optimal configuration of endowments, which results in the largest societal benefit, relies not just on inequality but on the unequal distribution of endowments to specifically favour "more productive individuals". In other words, the term productivity is used to mean 'effect of donation to public goods' but seems designed to imply 'productive' in its conventional sense.

The inference is that inequality is a path to optimality, whereas productivity is intrinsic and not related to individuals' endowments. Such ambiguous use of terminology risks compromising political impartiality and the goals of social equality and welfare.

**Stephen Thornquist** *F.M. Kirby Neurobiology Center, Boston, Massachusetts, USA. thornquist@fas.harvard.edu

## Anonymity calls for extreme caution

Confusion over data anonymization and privacy can have serious consequences when sensitive medical data are being collected for research. Anonymity cannot be achieved merely by dispensing with direct identifiers (see N. Seeman *Nature* **573**, 34; 2019).

People are identifiable in large data sets even in the absence of personal information (L. Sweeney *J. Law Med. Ethics* **25**, 98–110; 1997). For example, a few attributes such as demographic information can uniquely identify 99.98% of US subjects in any dataset (L. Rocher *et al. Nature Commun.* **10**, 3069; 2019). That is why recital 26 of the European Union's General Data Protection Regulation and section 1798.140 (h) of the California Consumer Privacy Act consider data as anonymous only when the subject cannot be re-identified.

Health research needs access to patient data to determine the precise patterns of signs and symptoms that indicate the onset of disease, and to monitor how these change in response to treatment. Because the mere absence of obvious identifiers does not protect privacy, it is imperative that such data continue to be collected, accessed and processed with caution and with strict security measures in place.

**Yves-Alexandre de Montjoye** *Imperial College London, UK.*
**Maxime Taquet** *University of Oxford, UK. demontjoye@imperial.ac.uk

## Don't pull punches in peer review

Holding reviewers to a code of conduct would be a mistake in my opinion, because it implies that the peer-review process should facilitate an author's research (see L. J. Beaumont *Nature* **572**, 439; 2019). Reviewers volunteer their time to judge the validity of a paper as a favour to the scientific community, not to the authors.

A code of conduct typically works best in situations that rely on volunteering and mentoring, where outcomes are not clear cut. For a research paper, this could preclude outright rejection by the reviewer, whose mandate would instead be to offer only constructive criticism to the authors. The role of a reviewer is to advise journal editors on a paper's suitability for publication, not to advise authors on how to make their work more acceptable to the journal. We already have mechanisms for providing some measure of constructive criticism — for example, when reviewers require major revisions.

Asking referees to keep their criticism positive could exacerbate the overall shortage of researchers willing to review manuscripts, particularly if they feel uncomfortable about reining in negative comments. The onus should instead be on the authors — to make their results clear and compelling in the first place.

**Rohit Goswami** *Indian Institute of Technology Kanpur, India. rgoswami@iitk.ac.in

**CONTRIBUTIONS**
Correspondence may be submitted to **correspondence@nature. com** after consulting the author guidelines and section policies at **go.nature.com/cmchno.**

# J. Robert Schrieffer

## *(1931–2019)*

### Physicist who shared Nobel for theoretical basis of superconductivity.

The story of how Robert Schrieffer solved a problem that had resisted the best minds in physics for more than 40 years, while riding the New York subway, is the stuff of legend in some circles. His explanation of how superconductivity works earned him a share of the 1972 Nobel Prize in Physics. A former president of the American Physical Society (APS), Schrieffer died on 27 July, aged 88.

In 1911 it was discovered that certain metals, when cooled to low enough temperatures, can carry current with no resistance. This seemingly miraculous property, superconductivity, arises directly from quantum mechanics, and underlies many contemporary technologies, such as magnetic resonance imaging body scanners and particle accelerators. For decades, however, there was no theory to explain how electrons in superconducting materials overcome their own mutually repulsive properties and other causes of resistance.

In early 1957, Schrieffer, then a 25-year-old graduate student, wrote down a quantum-mechanical wave function that accounted for the behaviour of electrons in superconductors. With his thesis adviser John Bardeen and postdoc colleague Leon Cooper, he published the now-famous BCS wave function and the full theory of superconductivity less than a year later — named BCS after the trio, who shared the Nobel prize (J. Bardeen, L. N. Cooper and J. R. Schrieffer *Phys. Rev.* **108**, 1175; 1957). The work has had far-reaching consequences for both fundamental science and practical technology. Schrieffer continued to make foundational contributions to our understanding of electrons in solids.

Born in Oak Park, Illinois, in 1931, Schrieffer studied physics at the Massachusetts Institute of Technology in Cambridge as an undergraduate. It was at graduate school at the University of Illinois at Urbana-Champaign that he began working with Bardeen, who in 1956 had just won a share of the physics Nobel for the invention of the transistor.

Bardeen suggested Schrieffer try his hand at understanding superconductivity. This was a risky proposition. After the initial success of quantum theory in describing ordinary conductors, insulators and semiconductors, there had been countless attempts to explain superconductors and all had failed. But the timing was right. Bardeen, with his then-postdoc David Pines,

had studied the effect of phonons (quantized sound waves) on metals, showing that they mediated an attractive interaction between electrons. Cooper found that this attractive interaction could lead to the formation of bound pairs of electrons. However, Cooper's theory described only the formation of a single electron pair. The question remained how to describe the many electrons pairing in the full electronic state of the metal, and why such pairing would lead to the properties of a superconductor.

Schrieffer's intuitive leap came to him on the subway while attending an APS meeting in 1957. It struck him that a natural wave function for describing a state with electron pairing was one in which the number of electrons was not fixed, but had a certain quantum mechanical uncertainty. He wrote it down there and then. This key insight, radical at the time but now part of the standard toolkit of theoretical physics, cracked the problem wide open. With the wave function in hand, it quickly became possible to calculate many of the observed properties of superconductors, and to predict new properties, which were subsequently found.

Schrieffer's beautiful idea has contributed to many branches of fundamental physics. In condensed-matter physics, it has also been applied to superfluid helium-3 and cold-atom systems. Elsewhere, the theory has helped to explain complex nuclei and neutron stars, and played a crucial part in establishing the understanding of quantum field theory that underlies today's standard model of strong, electromagnetic and weak interactions.

Schrieffer went on to take postdoctoral positions at the Niels Bohr Institute in Copenhagen and at the University of

Birmingham, UK. He held faculty positions at the University of Chicago, the University of Illinois and the University of Pennsylvania.

Throughout his career, Schrieffer displayed the same flair as in his brilliant wave function insight. In 1979, he and his colleagues showed that certain conducting polymers could exhibit excitations with electrical charge, but no spin (the magnetic moment of each electron is called its spin). The opposite could also occur: excitations could have spin, but no charge. It was a revelation that the two fundamental properties of electrons, charge and spin, could be split apart. This deconstruction has since been discovered at many other frontiers of condensed-matter physics. A later collaboration showed that a second example of deconstructed electrons, the fractionally charged excitations in the fractional quantum Hall states, also exhibit fractional statistics, meaning that they are not the conventional bosons or fermions that were thought to divide all fundamental particles into two classes.

In 1980, he moved to the University of California, Santa Barbara, and joined the newly formed Institute for Theoretical Physics. Here, between 1984 and 1989, he served as its second director, helping to establish its strong reputation as a centre for theoretical physics research. His final move in 1992 was back to Florida, where he took a state-wide professorial position in the Florida State University System. From that year until 2006 he was the first chief scientist of the National High Magnetic Field Laboratory at Florida State University in Tallahassee, where he had a crucial role in establishing the new facility's scientific credentials. His 1996 APS presidency was marked by his efforts to improve communication between the physics community and the public, and between physicists themselves to help unify the field.

Schrieffer was equally known for his warmth, charm, generosity and brilliance. When Bob discussed physics, his eyes would twinkle and a boyish demeanour would shine through. This enthusiasm and provision of wise counsel to younger physicists never waned. His unique style is captured, as if in a photograph, by the BCS wave function. ∎

*Nick Bonesteel* and *Gregory Boebinger* *are professors of physics who were colleagues of Schrieffer's at the National High Magnetic Field Laboratory and Florida State University.* *e-mails: bonestee@magnet.fsu.edu; gsb@magnet.fsu.edu*

BETTMAN/GETTY

# NEWS & VIEWS

# Getting to grips with bird landing

**Tree–dwelling birds can land on perches that vary in size and texture. Force measurements and video–footage analysis now reveal that birds rely on rapid and robust adjustments of their toe pads and claws to land stably.**

ANDREW A. BIEWENER

Even casual observations of flying birds, bats and insects reveal the adept and seemingly effortless ability of these creatures to land and take off safely from a wide variety of surfaces, whether these are tree branches, telephone wires, flowers or rocks. By contrast, passenger aircraft usually require long, flat runways to accomplish the same feats, and, even so, accidents can occur during take-off or landing. With the rise in the use of aerial drones for a range of applications[1-4], and the challenge of improving the aerodynamics and energy efficiency of drones, given their small size[5], there is interest in developing drone design to boost their success in landing on a range of complex surfaces. Writing in *eLife*, Roderick *et al.*[6] report their analysis of how Pacific parrotlets (*Forpus coelestis*) land on different types of perch, providing insights into the landing approach taken by these birds.

Previous work[7] has examined how vertebrates such as birds, bats and terrestrial mammals grip surfaces, by studying their feet and claws. This work has relied mainly on approaches such as comparative morphological analyses to assess foot, toe and claw geometry, studies of animal motion (termed descriptive kinematics) or static tests of grip strength. Such methods have shown, for example, how claw shape varies depending on an animal's size and claw use during its usual patterns of movement in its natural surroundings. For example, claws that are commonly used for running on the ground and manoeuvring usually have greater depth and are less curved than claws typically used for climbing. However, what has been lacking are studies of the dynamics and the forces that enable an animal to use its feet and claws to establish a stable support on landing, such as when birds perch.

Pacific parrotlets are tree-dwelling birds native to mountain forests of Ecuador and Peru. Roderick *et al.* studied how these birds landed (Fig. 1) on seven natural or artificial perches of differing diameters and textures, including rough, soft and slippery surfaces. Branches of three types of tree were tested, including one called a silk floss (*Ceiba speciosa*), found in the birds' natural habitat.

To independently monitor the front and rear
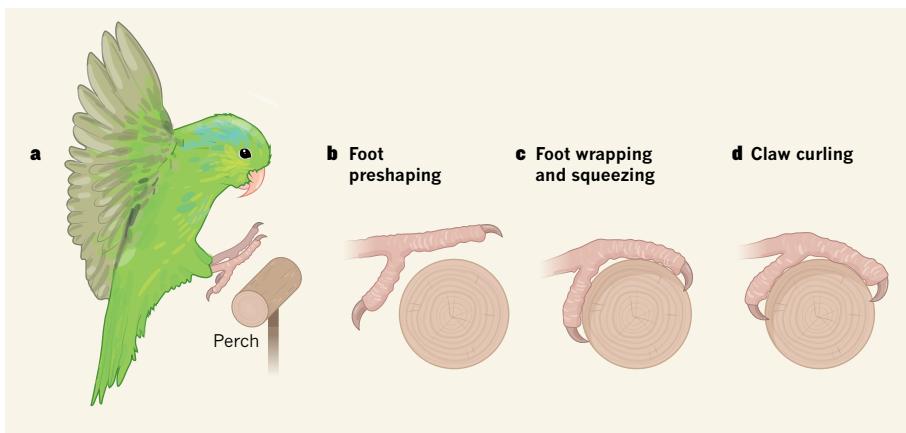


**Figure 1 | How a Pacific parrotlet (*Forpus coelestis*) lands stably on a perch.** Roderick *et al.*[6] analysed perching using methods to assess the forces that a bird encounters during landing, and by studying high-speed video recordings. **a**, When a bird is about to land, its wings, body and legs are positioned in the same, predictable way, consistent with earlier work[8,9] suggesting that birds use visual cues to position themselves for landing. At this stage, the bird's toes and claws are outstretched. **b**, When the bird is on the verge of making direct contact with the perch, its toes begin to close, in an event described as preshaping. **c**, When the bird's toes make contact with the perch, they wrap around it and squeeze it. **d**, The claws then begin to curl. This event can be superfast (1–2 milliseconds) if the perch surface is slippery.

of the landing surface of a perch, the authors designed split perches so that each half was anchored separately to a force and torque sensor that recorded the timing and features of the landing force and the rotational force experienced by the birds; both forces are influenced by the landing approach. The authors also measured the squeezing forces produced by the birds' feet and claws on landing. Combining these measurements with close-up, high-speed video recordings of the landing movements of the bird's wings, body, legs, feet and claws provided detailed information about the landing events associated with achieving a stable perch (see videos from the paper at go.nature.com/2nbfhtq and go.nature.com/2perfs9).

The authors report that the birds approached their landing on any given perch in a consistent fashion in terms of the movements of their wings and legs, with the landing and rotational forces varying uniformly over the time frame of each landing process. Such a landing strategy is consistent with earlier work[8,9] indicating that birds and insects approach a landing target using visual cues to accurately position their body appropriately for the estimated time when they will make contact with the landing surface.

This initial predictive phase of landing is followed by a rapid adjustment phase. It probably involves what is termed proprioceptive feedback from sensors in the bird's skin, muscles and joints, and communication with the nervous system, as the bird squeezes the perch, dragging its toe pads and claws across the perch's surface to achieve a stable grasp. Using laser scans and indentation tests to assess changes in the properties of the perch surface, Roderick and colleagues could relate the friction experienced by the birds' toes and claws to the animals' gripping movements, and showed how the movements of the bird's claws are adjusted to anchor the claws to perches of differing diameters and surface features.

The birds curled their claws more on perch surfaces that were difficult to grasp, such as those of large diameter or that generated low friction on landing, than on easier perches. During this grasping phase, the friction forces experienced by the toes (which are fairly consistent for a given perch type) are subsequently reinforced and are accompanied by less predictable, but higher gripping forces exerted on the perch surface by the claw tip. This strategy provides a stable safety margin for gripping the perch that is comparable to analogous safety

margins achieved by snakes[10] and robots[11], and is greater than the safety margins used by humans to grasp small objects[12]. Once stabilized on the perch, birds relax their grip, avoiding the unnecessary continued energy cost of muscle activation.

A limitation of Roderick and colleagues' work is that it did not investigate the role of the nervous system in controlling how gripping establishes a stable landing. The authors report superfast (1–2 milliseconds) initial anchoring movements of the claws, which suggests that these might be rapid, intrinsic, elastic mechanisms that do not involve neural control. However, these superfast movements are followed by longer-lasting adjustments in toe and claw movements that probably help to establish the stable grasp allowing birds to then relax their grip. These slower adjustments probably require proprioceptive feedback through the nervous system. Such feedback control could be evaluated by recording muscle activation and force patterns over the course of landing and perching. Inhibiting the activity of the

mechanosensory receptors in a bird's toe pads with an anaesthetic would offer a way to determine whether the loss of sensory feedback from toe pads affects these foot movements and the bird's landing ability.

The landing flights in this study were short and were made between perches on the same horizontal level. However, Pacific parrotlets probably fly to perches above or below the animal's current location when foraging. It would therefore be interesting to examine whether body orientation and landing forces vary depending on the trajectory of landing flights. Perhaps such flights might show less consistent patterns in the early stages of the landing process than were found by the authors. Nevertheless, Roderick and colleagues' detailed biomechanical analysis provides an important road map for future work on how feet, toes and claws enable animals to grip surfaces stably. ∎

**Andrew A. Biewener** *is in the Department of Organismic and Evolutionary Biology,*

*Harvard University, Massachusetts 01730, USA.*
*e-mail: abiewener@oeb.harvard.edu*

1. Cory, R. & Tedrake, R. in *AIAA Guidance, Navigation and Control Conf. Exhibit* https://doi.org/10.2514/6.2008-7256 (2008).
2. Pope, M. T. & Cutkosky, M. R. in *Biomimetic and Biohybrid Systems* (eds Lepora, N. F. *et al.*) 288–296 (Springer, 2016).
3. Kalantari, A., Mahajan, K., Ruffatto, D. & Spenko, M. in *2015 IEEE Int. Conf. Robot. Automation* 4669–4674 (IEEE, 2015).
4. Desbiens, A. L., Asbeck, A. T. & Cutkosky, M. R. *Int. J. Robot. Res.* **30**, 355–370 (2011).
5. Tennekes, H. *The Simple Science of Flight: From Insects to Jumbo Jets* (MIT Press, 2009).
6. Roderick, W. R. T., Chin, D. D., Cutkosky, M. R. & Lentink, D. *eLife* **8**, e46415 (2019).
7. Sustaita, D. *et al. Biol. Rev.* **88**, 380–405 (2013).
8. Lee, D. N., Davies, M. N. O., Green, P. R. & Van Der Weel, F. R. *J. Exp. Biol.* **180**, 85–104 (1993).
9. Wagner, H. *Nature* **297**, 147–148 (1982).
10. Byrnes, G. & Jayne, B. C. *Biol. Lett.* **10**, 20140434 (2014).
11. Estrada, M. A., Hawkes, E. W., Christensen, D. L. & Cutkosky, M. R. *2014 IEEE Int. Conf. Robot. Automation* 4215–4221 (IEEE, 2014).
12. Westling, G. & Johansson, R. S. *Exp. Brain Res.* **53**, 277–284 (1984).

---

# Dissolving without mixing

**Microfluidic devices have revolutionized biological assays, but complex set-ups are required to prevent the unwanted mixing of reagents in the liquid samples being analysed. A simpler solution has just been found.** SEE LETTER P.228

**ROBERT HOŁYST & PIOTR GARSTECKI**

On page 228, Gökçe *et al.*[1] report a clever solution to a fundamental problem in microfluidics: a simple and inexpensive method for delivering a liquid to multiple dried reagents that doesn't mix all the reagents together. By considering diffusion, convection (the flow along a channel) and capillary forces, the authors designed a microfluidic structure that produces a complicated, yet highly reproducible, liquid flow that first passes around dried spots of reagents and then back over them. This dissolves the dried reagents, but minimizes unwanted dispersal within the flow.

The 1990s saw an explosion of interest in microfluidics, driven by a vision of liquid-handling systems that were faster, simpler and smaller than existing devices being used in chemistry and biology. The fluid dynamics of liquids in microfluidic channels is fascinating: streams of distinct liquids typically flow side by side without turbulence or mixing[2], unlike liquid flows at larger scales. Convection in these systems can be tuned to rates similar to those of diffusion, which opens up a way to control the concentration gradients

of chemical reagents across parallel streams. Surprisingly, it was also found that the flow of immiscible liquids, which involves highly complex surface-tension forces, produces regular patterns of equally sized microdroplets in microchannels[3].

The ratio of the surface area of a microchannel-confined liquid (that is, the surface area bounded by the channel walls) to its volume is large, allowing heat and mass to be rapidly transferred to such liquids. Moreover, the flow of the liquid can be tightly controlled. Taken together, these features make microfluidics devices a useful platform for studying chemical reactions and biological processes. For example, miniature water droplets suspended in an oily continuous phase in microchannels can be used as reactors for chemical or biological processes.

The advent of microfluidics and droplet technologies led to breakthroughs in the life sciences. For example, these technologies have enabled digital assays[4] that can measure the concentration of specific genes in a sample without calibration. They are also key to the single-cell genetic-sequencing techniques[5] currently used in the Human Cell Atlas, a

project that aims to characterize every cell type in the human body[6]. Furthermore, microfluidics technologies are powering a wave of new point-of-care systems that bring diagnostic assays closer to the patient's bedside[7].

But a fundamental problem remains. In most applications, the microfluidic assay must run multiple analytical reactions on the same liquid sample. Each reaction requires a different reagent, which is dried and pre-stored on the cartridge before the sample is added. These reagents should not mix with each other, because this would ruin the assay. But mixing is hard to avoid once the sample has been added, because of dispersion effects in the liquid. Several solutions to this problem have been proposed, always involving two steps — one to deliver the sample to the reagents, and the other to isolate the microchambers in which the reagents are stored from each other. The second step typically either uses an immiscible liquid as a barrier, or the microchambers are enclosed by solid walls, but either option complicates the design, manufacturing and use of these systems.

Gökçe *et al.* have tackled the problem in a much simpler way. They prepared a straight section of channel that is divided into two along its length by a shallow barrier, and deposited dried spots of reagents in one of the resulting halves (Fig. 1). They then introduced a sample liquid so that it filled the other half of the channel, before changing direction to bend around the end of the barrier and fill the portion of the channel containing the dried spots. Once the whole channel has been filled, the resulting solution of reagents is released through a valve so that it can enter the next section of the microfluidic system. This produces a solution that has an approximately uniform concentration of reagents throughout
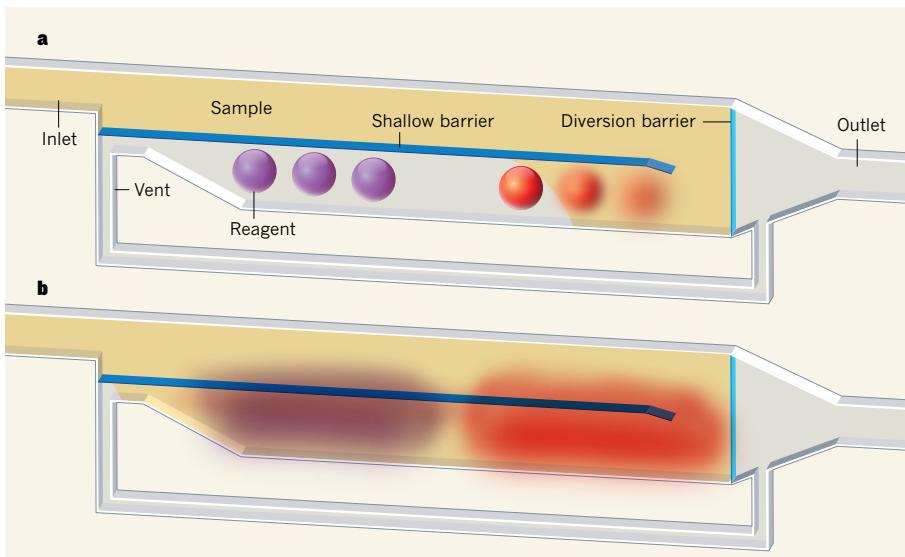
**Figure 1 | A module for microfluidics. a**, In Gökçe and colleagues' microfluidic architecture[1], a straight microchannel is divided by a shallow barrier, and dried reagents are spotted along one half. A liquid sample entering from the inlet first passes down one side, and then fills the side containing the reagents. Air pushed ahead of the moving front of the liquid escapes through a vent. **b**, The dried reagents dissolve, and the resulting solution spills over the shallow barrier to fill the whole channel. The capillary forces generated in the system prevent the reagents from being dispersed so that they become concentrated at the moving front, as they would have been in a simple channel. Once the channel is full, the liquid is released through the diversion barrier to the outlet. The system allows multiple reagents to be dissolved in a liquid sample without being mixed together by dispersion.

its volume. By contrast, when dried reagents are dissolved by a liquid in a simple, unstructured microchannel, dispersion processes cause the reagents to become concentrated at the moving front of the liquid.

The authors went on to demonstrate how their system could be used to precisely control the concentration and the timing of addition of reagents in complex biochemical reactions, in two assays: one that detected DNA sequences of the human papilloma virus, and the other that quantified the activity of an enzyme. In both cases, the assays involved the use of several reagents (enzymes and their substrates, cofactors, fluorescent reporter molecules, and so on).

The key to Gökçe and co-workers' invention is the shallow barrier in the channel, which acts as a capillary pinning line — an interface with the liquid that constrains the liquid's motion through capillary forces. The phenomenon of capillary pinning is common in nature; for example, it holds water droplets to minuscule specks of dirt on glass. Capillary-pinning lines underlie such diverse effects as the formation of coffee rings from droplets spilt on a table[8], or the unidirectional flow of water in the carnivorous pitcher plant *Nepenthes alata*[9].

Capillary pinning has been used in microfluidics systems before, for example in capillary valves[10], which control liquid flow without using mechanical parts. They have also been used in phaseguides, which form barriers to flow perpendicular to the direction of motion of the liquid–air meniscus — these barriers hold the meniscus until enough

pressure has built up for liquid to flow over the barrier[11]. Gökçe *et al.* have used capillary pinning in a new way: to enable liquids to flow over dried spots of reagents without causing the reagents to disperse uncontrollably within the liquid, thus allowing the concentration profile of the reagents in the resulting

solution to be controlled by the positioning of the original spots.

The authors' use of small-scale capillary forces allowed them to segregate reactions without using solid walls. This opens up a simple approach for preprogramming and implementing large numbers of biochemical reactions in straight microchannels, removing the need for complex microfluidic chips that have large numbers of compartments and valves. The authors also show that the geometries of their microchannel systems can be made using inexpensive mass-production methods. These systems could therefore help to bring increasingly sophisticated biochemical assays closer to patients in point-of-care devices. ∎

**Robert Hołyst** *and* **Piotr Garstecki** *are at the Institute of Physical Chemistry, Polish Academy of Sciences, Warsaw 01-224, Poland. e-mails: rholyst@ichf.edu.pl; garst@ichf.edu.pl*

1. Gökçe, O. *et al. Nature* **574**, 228–232 (2019).
2. Kenis, P. J. A., Ismagilov, R. F. & Whitesides, G. M. *Science* **285**, 83–85 (1999).
3. Teh, S.-Y., Lin, R., Hung, L.-H. & Lee, A. P. *Lab Chip* **8**, 198–220 (2008).
4. Hindson, C. M. *et al. Nature Methods* **10**, 1003–1005 (2013).
5. Prakadan, S. M., Shalek, A. K. & Weitz, D. A. *Nature Rev. Genet.* **18**, 345–361 (2017).
6. Hon, C.-C., Shin, J. W., Carninci, P. & Stubbington, M. J. T. *Briefings Funct. Genom.* **17**, 283–294 (2018).
7. Chin, C. D., Linder, V. & Sia, S. K. *Lab Chip* **12**, 2118–2134 (2012).
8. Deegan, R. D. *et al. Nature* **389**, 827–829 (1997).
9. Chen, H. *et al. Nature* **532**, 85–89 (2016).
10. Glière, A. & Delattre, C. *Sensors Actuators A* **130–131**, 601–608 (2006).
11. Vulto, P. *et al. Lab Chip* **11**, 1596–1602 (2011).

# Early European babies bottle–fed animal milk

**The foods used to supplement or replace breast milk in infants' diets in prehistoric times aren't fully understood. The finding that ancient feeding vessels from Europe had residues of animal milk offers a clue.** SEE LETTER P.246

**SIÂN E. HALCROW**

Small pottery vessels, sometimes with animal-like forms (Fig. 1), containing a spout through which liquid could be poured, have been found at prehistoric archaeological sites in Europe. One idea put forward is that they were used as feeding vessels for sick adults and the elderly. However, on page 246, Dunne *et al.*[1] describe an analysis of spouted vessels found in ancient graves of infants in Germany that indicates that these artefacts contained animal milk. This evidence suggests

that such vessels were used to feed animal milk to children, providing crucial insight into the diet of developing infants in prehistoric human populations.

For years, many archaeologists ignored children when studying ancient populations, but researchers now increasingly recognize the importance of children when trying to understand the factors affecting earlier societies[2,3]. One such example concerns a major societal turning point in human prehistory, known as the Neolithic demographic transition, when there is evidence of a substantial increase

in fertility and a growth in the number of individuals in human populations compared with that of earlier societies[4].

The Neolithic period in Europe began roughly around 7000 BC. During the Neolithic, some humans began to move away from a hunter-gatherer lifestyle towards one that depended on crops and domesticated animals. How did this transition to agriculture lead to a baby boom? An exploration of the approaches used to feed infants might provide some of the evidence needed to answer this question.

Some of the earliest known pottery vessels of a suitable size and shape for use in feeding infants are from the Neolithic period. These artefacts, discovered in Germany, have been dated[5] to between 5500 and 4800 BC. It has been suggested[6] that during the Neolithic, weaning — when an infant's diet changes from breast milk to other foods — occurred earlier in an infant's life than was previously the case. This earlier weaning might have been accomplished by using animal milk and plant sources of carbohydrates. It has been argued that such early weaning could have helped to counteract the period of infertility that can occur while a mother is breastfeeding[7], and thus might have led to the increase in fertility and population size during the Neolithic demographic transition. In the archaeological record, this fertility increase is evidenced, somewhat counter-intuitively, by an increase in the number of infants found at burial sites — if more babies are born in a population, then more babies will also die, and be buried[8].

Dunne and colleagues examined ceramic vessels with spouts found in children's graves from burial sites in Bavaria, Germany. One vessel came from a burial site dated to around 1200–800 BC (during the late Bronze Age), and two vessels came from a burial site from around 800–450 BC (during the early Iron Age).

The authors analysed traces of ancient food in these vessels to determine the origin of these residues, by assessing specific characteristics of fatty-acid molecules. Dunne *et al.* used isotope analysis to study the chemistry of specific compounds in the vessels, and also obtained molecular 'fingerprints' of the ancient lipids. They then compared this information with the fingerprints of known reference compounds. This evidence indicates that the vessels contained fatty acids from dairy products, probably milk, that came from domestic ruminant animals. The specific type of animal that provided this milk was not identified.

It is thought that humans first started drinking animal milk in Europe. A study[9] published this year of proteins captured in dental plaque provides direct evidence that adults drank animal milk during the Neolithic period in Europe, with the earliest dates for this occurring around 6,000 years ago. Now Dunne *et al.* present the earliest known evidence of animal milk in small bottles for infants.

The exploration of infant feeding provides



**Figure 1 | Ancient pottery vessels.** Vessels with a spout for pouring liquid and of a size suitable for feeding babies have been found at archaeological sites. The earliest examples of such vessels[5] have been dated to around 5500–4800 BC, but whether these were used to feed infants is unknown. Two vessels are shown of this size and shape from the late Bronze Age or early Iron Age (vessels dated between 1200 and 800 BC). The vessel on the left, from Vösendorf, Austria, is approximately 90 millimetres high. The vessel on the right, from Statzendorf, Austria, is about 85 mm high. Dunne and colleagues' analysis[1] of organic residues found in ancient spouted vessels (not those pictured) sheds light on how early populations might have fed young infants.

information about how babies have been cared for and how social attitudes towards infant feeding have changed over time[10]. Dunne and colleagues' investigation of infant feeding during the Neolithic provides insight into cultural beliefs related to the body, infancy and motherhood. Furthermore, the type of food infants are fed, and when during their development they are given food in addition to breast milk, has a strong relationship to infant health and survival[11].

Human breast milk is a perfect baby food, containing carbohydrates, protein, fat, vitamins, minerals, digestive enzymes and hormones[12]. It provides protection from infection because it contains numerous types of immune cell[13,14]. Some of the sugars it contains, although not digested by babies, support certain communities of gut microorganisms, which prevent disease-causing microbes from establishing a presence in the body[14]. By contrast, animal-milk products do not provide a complete nutritional source for infants. And the use of hard-to-clean bottles for animal milk poses a risk of exposure to life-threatening infections such as gastroenteritis. The introduction of milk in bottles during the Neolithic, therefore, might have led to a deterioration in the health of some infants.

Further research on the remains of people in European prehistoric cemetery sites should be undertaken to explore the effects of the introduction of animal milk as an infant food. This could be assessed by analysing the rate of infant and child mortality, and determining whether any signs of nutritional or infectious disease are present when studying the bones and teeth in infant remains. Furthermore, the age at which a child was weaned can be investigated using techniques that analyse teeth[15], and gathering such data can uncover the variation in weaning approaches that existed

in a population[16]. Such knowledge, together with evidence of disease for the individual being studied, might help to provide a greater understanding of the significance of the introduction of animal milk for the lives of ancient children. ∎

**Siân E. Halcrow** *is in the Department of Anatomy, University of Otago, Dunedin 9016, New Zealand.*
*e-mail: sian.halcrow@otago.ac.nz*

1. Dunne, J. *et al.* *Nature* **574**, 246–248 (2019).
2. Lillehammer, G. in *The Oxford Handbook of the Archaeology of Childhood* (eds Crawford, S., Hadley, D. & Shepherd, G.) 38–51 (Oxford Univ. Press, 2018).
3. Halcrow, S. & Tayles, N. in *Social Bioarchaeology* (eds Agarwal, S. & Glencross, B.) 333–360 (Oxford Univ. Press, 2011).
4. Bocquet-Appel, J. P. *Science* **333**, 560–561 (2011).
5. Meller, H. E. *Bronzerausch: Spätneolithikum und Frühbronzezeit. Begleithefte zur Dauerausstellung 4* (State Museum of Prehistory Halle, 2011).
6. Bocquet-Appel, J.-P. *Curr. Anthropol.* **43**, 637–650 (2002).
7. Chao, S. *Clin. Perinatol.* **14**, 39–50 (1987).
8. Jackes, M. in *Strength in Diversity: A Reader in Physical Anthropology* (eds Herring, A. & Chan, L.) 155–185 (Canadian Scholars, 1994).
9. Charlton, S. *et al.* *Archaeol. Anthropol. Sci.* https://doi.org/10.1007/s12520-019-00911-7 (2019).
10. Tomori, C., Palmquist, A. E. L. & Quinn, E. A. in *Breastfeeding: New Anthropological Approaches* (eds Tomori, C., Palmquist, A. E. L. & Quinn, E. A.) Ch. 1, 1–25 (Routledge, 2017).
11. Halcrow, S. E. *et al.* in *Breastfeeding: New Anthropological Approaches* (eds Tomori, C., Palmquist, A. E. L. & Quinn, E. A.) Ch. 11, 155–169 (Routledge, 2017).
12. Lessen, R. & Kavanagh, K. J. *Acad. Nutr. Diet.* **115**, 444–449 (2015).
13. Martin, C. R., Ling, P.-R. & Blackburn, G. L. *Nutrients* **8**, 279 (2016).
14. Allens-Blevins, C. R., Sela, D. A. & Hinde, K. *Evol. Med. Public Health* **2015**, 106–121 (2015).
15. Beaumont, J., Gledhill, A., Lee-Thorp, J. & Montgomery, J. *Archaeometry* **55**, 277–295 (2013).
16. King, C. L. *et al.* *Int. J. Osteoarchaeol.* **28**, 599–612 (2018).

TUMOUR BIOLOGY

# Fungi accelerate pancreatic cancer

**The impact of fungi on human health is under-studied and underappreciated. One genus of fungus, *Malassezia*, has now been linked to the progression of pancreatic cancer.** SEE LETTER P.264

**IVY M. DAMBUZA & GORDON D. BROWN**

The communities of microorganisms that occupy specific regions of the body are often altered in cancer[1], and these microbiomes — particularly their bacterial components — are a current focus of cancer research. One example is pancreatic ductal adenocarcinoma (PDA), for which changes in the bacterial community occupying the pancreas have been documented[2]. This lethal disease often goes undetected until it has reached advanced stages, and the prognosis is usually very poor[3]. Aykut *et al.*[4] reveal on page 264 that the fungal component of the pancreatic microbiome (known as the mycobiome) is also altered in PDA. In fact, an abundance of a specific fungal genus actually promotes the disease.

The mycobiome is a historically under-recognized player in human health and disease, but its role in both is essential. Harmless organisms called commensals, including fungi, inhabit mucosal surfaces such as the linings of the gut, nose and mouth, and can activate inflammatory processes as part of the immune system's response to injury or infection. In some cases, changes in the biodiversity of fungal communities are linked to aggravated inflammatory-disease outcomes.

For example, intestinal overgrowth of *Candida albicans* — a fungus that causes oral thrush in babies — has been associated with severe forms of intestinal ulcers[5] and with mould-induced asthma[6]. Moreover, it is becoming apparent that there is a relationship between the gut mycobiome and human cancers, including colorectal and oesophageal cancer[7].

Aykut *et al.* used DNA sequencing to search for fungus-specific genomic markers in the cancerous pancreas. This revealed increased pancreatic fungal colonization, both in humans who have PDA and in experimental mouse models of PDA, compared with the pancreas of healthy counterparts. What is the source of these fungi? The authors introduced a fluorescently tagged fungal strain into the guts of mice, and the fungus could be detected in the pancreas as early as 30 minutes later. It is known that there is a direct link between the gut and the pancreatic duct, and microbial translocation into the pancreas has been seen for other organisms[8], but not previously for fungi.

> *"The mycobiome is a historically under-recognized player in human health and disease."*

The researchers then investigated the link between pancreatic tumour development and fungi using mice engineered to express a cancer-causing protein in the pancreas. These mice develop a slowly progressive PDA that recapitulates the human disease. The mycobiome of the pancreas was notably different from that of the gut in the mutant mice, although the mechanisms underlying this difference are unclear. One genus of yeast, *Malassezia*, was much more prevalent in pancreatic tumours than in either the guts of these animals or the pancreas of healthy animals. Importantly, *Malassezia* was also prevalent in human PDA samples.

*Malassezia* species have been best studied in skin conditions such as dandruff and atopic dermatitis. Indeed, they are the most abundant fungal species in mammalian skin, accounting for more than 80–90% of the skin's commensal mycobiome[9]. Because we are constantly exposed to *Malassezia*, healthy individuals can have immune responses to the genus, which in some cases lead to disease. For instance, inflammation caused by overgrowth of *Malassezia* can worsen gastric ulcers[10].

This information hinted that the abundance of *Malassezia* in PDA tumours could be medically relevant. Indeed, Aykut *et al.* found that antifungal drugs halted PDA progression in mice, and improved the ability of chemotherapy to shrink the tumour. Subsequent repopulation of the antifungal-treated animals with a *Malassezia* species accelerated PDA growth again.

Next, Aykut and colleagues asked how *Malassezia* promotes PDA growth. Gene-expression analysis revealed that poor survival outcome in human PDA was associated with expression of a molecule called mannose binding lectin (MBL).

MBL is a soluble protein produced in the liver that binds carbohydrates on the surface of microorganisms and then activates a protein system called the complement cascade in the blood. The complement cascade serves a variety of immune functions, including activating immune cells to ingest and kill fungi and other pathogens. The cascade has also been linked to tumour development, because its pro-inflammatory pathways stimulate the growth, survival and motility of cells — including cancer cells. In a final set of experiments, Aykut *et al.* found that PDA progression was delayed in mice lacking MBL or a key component of the complement cascade called C3, even if *Malassezia* was present in the pancreas. Thus, *Malassezia* augments PDA progression by promoting pancreatic inflammation through the complement cascade (Fig. 1).

Aykut and colleagues' results reveal a previously unappreciated role for fungi in PDA progression. A valuable next step will be to determine whether this role somehow involves interactions with the bacterial species known to promote PDA progression[3]. Fungi and bacteria coexist in the gut and other mucosal
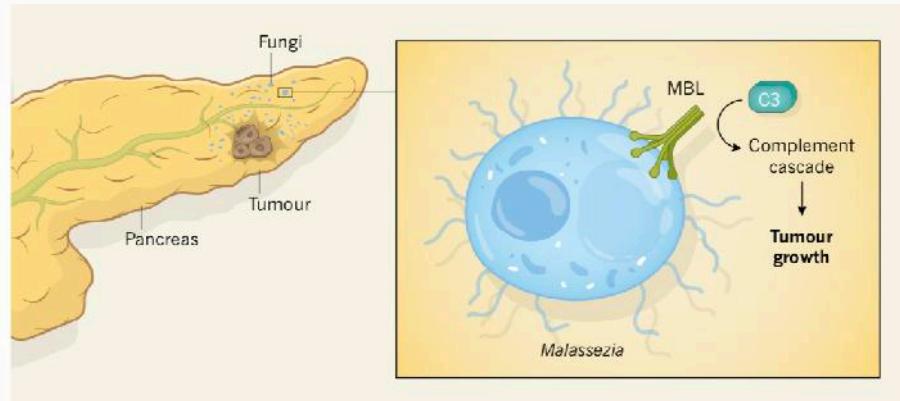
**Figure 1 | Fungi called *Malassezia* promote pancreatic ductal adenocarcinoma.** Aykut *et al.*[4] report that the community of fungi that inhabits the pancreas is altered when mice or humans have the cancer pancreatic ductal adenocarcinoma (PDA), with species of the genus *Malassezia* becoming particularly abundant. The extracellular protein mannose binding lectin (MBL) recognizes an unidentified carbohydrate structure expressed by *Malassezia* and activates the protein C3, triggering an inflammatory immune response called the complement cascade. Complement activation has many effects, including stimulation of cell growth, survival and migration — factors that fuel tumour growth.

sites, and it is likely that alterations in one community will affect the other. In some scenarios, disease-specific coexistence of bacteria and fungi has been noted — for instance, bacteria of the genus *Pseudomonas* are often isolated from the lungs of people with cystic fibrosis, which are often infected with fungi called *Aspergillus*[10]. Understanding these microbial networks will further enhance our understanding of disease progression and inform therapeutic interventions.

Another unresolved question is how MBL and the complement system integrate with the rest of the immune system during PDA progression. For example, how do MBL and the complement cascade interact with the signalling pathways triggered by an immune-cell receptor protein called dectin-1? This protein recognizes the fungal cell wall and activates protective antifungal immune pathways, often in collaboration with other receptors, including those that recognize the complement cascade. In addition, dectin-1 can directly recognize proteins on tumour cells and modulate the activity of tumour-killing immune cells[11]. But dectin-1 can also associate with tumour-recognizing receptors, which can promote PDA progression[12]. Thus, it is clear that we need a much better understanding of the complex interplay between the components of the immune system that target fungi and those that target tumours.

This study highlights a role for fungi in the development of cancer. Excitingly, the work points to the possibility of new therapeutic approaches. Perhaps altering microbial communities by directly targeting specific populations could help ameliorate PDA. Alternatively, therapies targeting immune components such as MBL that control fungal infections could provide a route to combat this lethal cancer. ∎

**Ivy M. Dambuza** *and* **Gordon D. Brown** *are in the Medical Research Council Centre for Medical Mycology, University of Exeter, Exeter EX4 4QD, UK.*
*e-mails: i.m.dambuza@exeter.ac.uk; gordon.brown@exeter.ac.uk*

1. Helmink, B. A. *et al. Nature Med.* **25**, 377–388 (2019).
2. Pushalkar, S. *et al. Cancer Discov.* **8**, 403–416 (2018).
3. McGuigan, A. *et al. World J. Gastroenterol.* **24**, 4846–4861 (2018).
4. Aykut, B. *et al. Nature* **574**, 264–267 (2019).
5. Iliev, I. D. *et al. Science* **8**, 1314–1317 (2012).
6. Bacher, B. *et al. Cell* **6**, 1340–1355.e15 (2019).
7. Elinav, E., Garrett, W. S., Trinchieri, G. & Wargo, J. *Nature Rev. Cancer* **7**, 371–376 (2019).
8. Schmid, St. W., Uhl, W. & Büchler, M. W. in *Pancreatic Disease: State of the Art and Future Aspects of Research* (eds Lankisch, P. G. & DiMagno, E. P.) 39–54 (Springer, 1999).
9. Findley, K. F. *et al. Nature* **498**, 367–370 (2013).
10. Kruger, W. *et al. Pathogens* **8**, 70 (2019).
11. Chiba, S. *et al. eLife* **3**, e04177 (2014).
12. Daley, D. *et al. Nature Med.* **23**, 556–567 (2017).

This article was published online on 2 October 2019.

# Predicting if the worst earthquake has passed

**When a big earthquake occurs, it is hard to tell if it will be followed by a larger quake or by only smaller ones. A method has been developed that aims to distinguish between these scenarios while events are still unfolding.** SEE ARTICLE P.193

**EMILY E. BRODSKY**

After every major earthquake, seismologists warn the public that the danger has not yet passed: aftershocks will continue to shake the ground. These aftershocks usually get smaller over time, but, occasionally, an aftershock will be larger than the original event. Standard earthquake statistics suggest that the latter situation should occur about 5–10% of the time[1,2], but is there any way of knowing which aftershock sequences will behave in this anomalous way? More simply, after a big earthquake, is it possible to determine whether an even larger one is coming? On page 193, Gulia and Wiemer[3] propose an answer to this question. They suggest that, by continuously measuring the relative numbers of large and small earthquakes, comparatively safe aftershock sequences can be distinguished from those that will get bigger.

The magnitude distribution of earthquakes generally follows a relationship known as the Gutenberg–Richter law[4]. Roughly speaking, in most places on Earth, for every earthquake of magnitude 4 or larger, there will be 10 quakes of magnitude 3 or larger and 100 quakes of magnitude 2 or larger. The exact ratio of big to small earthquakes in a particular time or place is described by a parameter called the *b* value. If this value is low, there will be comparatively fewer small quakes for every big one. And if it is high, there will be more small quakes for every big one.

In previous work, Gulia and Wiemer, together with co-workers, found that the

**Figure 1 | Damage caused by an earthquake aftershock in Norcia, Italy.** On 30 October 2016, the town of Norcia was hit by the aftershock of a large earthquake that had occurred two months previously. Unlike most aftershocks, this one was bigger than the original quake. Gulia and Wiemer[3] suggest that it might be possible to predict whether a large earthquake will be followed by a bigger aftershock or by only smaller ones.

$b$ value normally rises during an aftershock sequence, which means that small earthquakes become more common[5]. In the present work, the authors noticed that, occasionally, the $b$ value drops instead of rising, implying that big quakes increase in frequency. They also noticed that these sequences are the only ones that contain an aftershock larger than the original quake.

According to the definition of the $b$ value, sequences that have low values are more likely to be associated with big earthquakes than are those that have high values. Therefore, Gulia and Wiemer's finding might seem to be merely a restatement of aftershock statistics. However, the authors suggest that the observed pattern is deterministic rather than statistical, on the basis of the fact that a falling $b$ value is seen robustly for only two earthquake sequences in the entire data set: the 2016 Kumamoto earthquakes in Japan and the 2016 Amatrice–Norcia earthquakes in Italy (Fig. 1). Each of these sequences contained an anomalously large and damaging aftershock. For nearly all of the other sequences, the $b$ value increased directly after the original quake. The authors note one exception to this, which they attribute to poor data quality in the early 1980s.

Making such a claim based on two aftershock sequences might seem bold. But in earthquake science, we are often driven to closely analyse the few examples that are available because nature provides only uncontrolled experiments at irregular intervals. Nonetheless, we need to proceed with extreme caution in the face of such sparse data.

In particular, measuring the magnitude distribution is not as simple as it at first seems. Many judgement calls are required to determine how big the measurement region should be, how to define the normal $b$ value for a region and how to account for the fact that many aftershocks are not recorded in the wake of a large earthquake. These decisions must be made for each region, and the decision-making is the Achilles heel of statistical seismology studies such as this one.

For instance, the authors opt to use data collected at least 3 days after the first large Amatrice–Norcia earthquake to compute the $b$ value, but used data collected at least 0.05 days after the first Kumamoto event, because of the higher quality of the Japanese earthquake catalogue. If they had waited 0.2 days after the first Kumamoto quake, their traffic-light coding system would have given a yellow warning rather than a red one — that is, a less-definitive warning.

Expert judgement is intrinsic to the design of scientific analyses and, in this case, a different judgement would have led to a different answer. So how can we determine whether the correct decisions have been made? The gold standard of any scientific theory is its ability to predict data that have not been collected when the theory is proposed. Gulia and Wiemer have documented their decisions through a full release of their computer code. As new earthquakes occur, the key test of the paper will be in the reuse of this code.

Earth is already providing us with opportunities to test the authors' claim. The 2019 Ridgecrest earthquakes in California are notable for having a magnitude-6.4 event followed within days by a magnitude-7.1 event (see go.nature.com/2pjalib). Other examples will surely follow. We can all hope for a more predictable future in which these anomalous events cease to be surprises. ∎

**Emily E. Brodsky** *is in the Department of Earth and Planetary Sciences, University of California, Santa Cruz, Santa Cruz, California 95064, USA.*
*e-mail: brodsky@ucsc.edu*

1. Jones, L. M. *Bull. Seismol. Soc. Am.* **75**, 1669–1679 (1985).
2. Felzer, K. R., Abercrombie, R. E. & Ekström, G. *Bull. Seismol. Soc. Am.* **94**, 88–98 (2004).
3. Gulia, L. & Wiemer, S. *Nature* **574**, 193–199 (2019).
4. Kanamori, H. & Brodsky, E. E. *Phys. Today* **54**, 34–40 (2001).
5. Gulia, L. *et al. Geophys. Res. Lett.* **45**, 13277–13287 (2018).

# The human body at cellular resolution: the NIH Human Biomolecular Atlas Program

HuBMAP Consortium*

**Transformative technologies are enabling the construction of three-dimensional maps of tissues with unprecedented spatial and molecular resolution. Over the next seven years, the NIH Common Fund Human Biomolecular Atlas Program (HuBMAP) intends to develop a widely accessible framework for comprehensively mapping the human body at single-cell resolution by supporting technology development, data acquisition, and detailed spatial mapping. HuBMAP will integrate its efforts with other funding agencies, programs, consortia, and the biomedical research community at large towards the shared vision of a comprehensive, accessible three-dimensional molecular and cellular atlas of the human body, in health and under various disease conditions.**

The human body is an incredible machine. Trillions of cells, organized across an array of spatial scales and a multitude of functional states, contribute to a symphony of physiology. While we broadly know how cells are organized in most tissues, a comprehensive understanding of the cellular and molecular states and interactive networks resident in the tissues and organs, from organizational and functional perspectives, is lacking. The specific three-dimensional organization of different cell types, together with the effects of cell–cell and cell–matrix interactions in their natural milieu, have a profound impact on normal function, natural ageing, tissue remodelling, and disease progression in different tissues and organs. Recently, new technologies have enabled the molecular characterization of a multitude of cell types[1–4] and mapping of their spatial relationships in complex tissues at unprecedented scale and single-cell resolution. These advances create the opportunity to build a high-resolution atlas of three-dimensional maps of human tissues and organs.

HuBMAP (https://commonfund.nih.gov/hubmap) is an NIH-sponsored program with the goals of developing an open framework and technologies for mapping the human body at cellular resolution as well as generating foundational maps for several tissues obtained from normal individuals across a wide range of ages. A previous NIH-sponsored project, GTEx[5], examined DNA variants and bulk tissue expression patterns across approximately a thousand individuals, but HuBMAP is a distinct project focused on generating molecular maps that are spatially resolved at the single-cell level but using samples from a more limited number of people. To achieve these goals, HuBMAP has been designed as a cohesive and collaborative organization, with a culture of openness and sharing using team science-based approaches[6]. The HuBMAP Consortium (https://hubmapconsortium.org/) will actively work with other ongoing initiatives including the Human Cell Atlas[7], Human Protein Atlas[8], LIfeTime (https://lifetime-fetflagship.eu/), and related NIH-funded consortia that are mapping specific organs (including the brain[9], lungs (https://www.lungmap.net/), kidney (https://kpmp.org/about-kpmp/), and genitourinary (https://www.gudmap.org/) regions) and tissues (especially pre-cancer and tumours[10]; https://humantumoratlas.org), as well as other emerging programs.

## HuBMAP organization and approaches

The HuBMAP consortium comprises members with diverse expertise (for example, molecular, cellular, developmental, and computational

biologists, measurement experts, clinicians, pathologists, anatomists, biomedical and software engineers, and computer and data information scientists) and is organized into three components: (1) tissue mapping centres (TMCs); (2) HuBMAP integration, visualization and engagement (HIVE) collaborative components; and (3) innovative technologies groups (transformative technology development (TTD) and rapid technology implementation (RTI)) (Fig. 1). Throughout the program, HuBMAP will increase the range of tissues and technologies studied through a series of funding opportunities that have been designed to be synergistic with other NIH-funded and international efforts. In the later stages of HuBMAP, demonstration projects will be added to show the utility of the generated resources and, importantly, to engage the wider research community to analyse HuBMAP data alongside data from other programs or their own labs.

## Tissue and data generation

The HuBMAP TMCs will collect and analyse a broad range of largely normal tissues, representing both sexes, different ethnicities and a variety of ages across the adult lifespan. These tissues (Fig. 2) include: (1) discrete, complex organs (kidney, ureter, bladder, lung, breast, small intestine and colon); (2) distributed organ systems (vasculature); and (3) systems comprising dynamic or motile cell types with distinct microenvironments (lymphatic organs: spleen, thymus, and lymph nodes). Tissue will be collected at precisely defined anatomical locations (when possible, photographically recorded) according to established protocols that preserve tissue quality and minimize degradation. Beyond meeting standard regulatory requirements, to the greatest extent possible, consent will be obtained so that the generated data is available for open-access data sharing (that is, public access without approval by data committees), to maximize their usage by the biomedical community.

To achieve spatially resolved, single-cell maps, the TMCs will use a complementary, iterative, two-step approach (Fig. 3). First, 'omic assays, which are extremely efficient for data acquisition, will be used to generate global genome sequence and gene expression profiles of dissociated single cells or nuclei in a massively parallel manner. The molecular state of each cell will be revealed by single-cell transcriptomic[11] and, in many cases, chromatin accessibility[12,13] assays; imputation of transcription

---
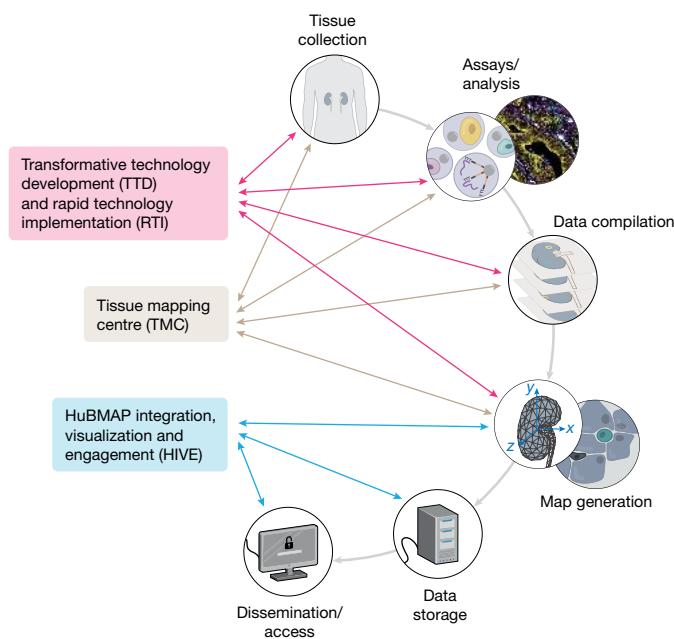
**Fig. 1 | The HubMAP consortium.** The TMCs will collect tissue samples and generate spatially resolved, single-cell data. Groups involved in TTD and RTI initiatives will develop emerging and more developed technologies, respectively; in later years, these will be implemented at scale. Data from all groups will be rendered useable for the biomedical community by the HuBMAP integration, visualization and engagement (HIVE) teams. The groups will collaborate closely to iteratively refine the atlas as it is gradually realized.



**Fig. 2 | Key tissues and organs initially analysed by the consortium.** Using innovative, production-grade ('shovel ready') technologies, HuBMAP TMCs will generate data for single-cell, three-dimensional maps of various human tissues. In parallel, TTD projects (and later RTI projects) will refine assays and analysis tools on a largely distinct set of human tissues. Samples from individuals of both sexes and different ages will be studied. The range of tissues will be expanded throughout the program.

factor binding regions from the open chromatin data combined with the gene expression data will be used to explain the regulation of gene expression across the distinct cell types[14]. Second, spatial information (abundance, identities, and localization) will be acquired for various biomolecules (RNA[15], protein[16], metabolites, and lipids) in tissue sections or blocks, using imaging methodologies such as fluorescent microscopy (confocal, multiphoton, lightsheet, and expansion), sequential fluorescence in situ hybridization (seqFISH)[17,18], imaging mass spectrometry[19,20], and imaging mass cytometry (IMC)[21–24]. The extensive single-cell and nucleus profiles obtained will inform in situ modalities (for example, single-cell or nucleus RNA sequencing will be used to choose probes for RNA or proteins), which will provide spatial information for up to hundreds of molecular targets of interest. These data will allow the computational registration of cell-specific epigenomic or transcriptomic profiles to cells on a histological slide to reveal various microenvironmental states. They will potentially include information about protein localization to cytoplasm, nucleus, or cell surface; phosphorylation; complex assembly; extracellular environment; and cellular phenotype determined by protein marker coexpression. Registration and computational integration of complex imaging data will provide biological insights beyond any single imaging mode[19,25]. The powerful combination of single-cell profiling and multiplexed in situ imaging will provide a pipeline for constructing multi-omics spatial maps for the various human organs and their cellular interactions at a molecular level.

The TMCs will use complementary methods for data collection with an emphasis on processes to ensure the generation of high-quality data and standardized metadata annotations. Benchmarking, quality assurance and control standards, and standard operating procedures, where appropriate, will be developed for each stage of the methodological process and be made available to promote rigor, reproducibility and transparency. It is expected that quality assurance and control standards for both biospecimens and data will evolve as tissue collection, processing techniques, storage and shipping conditions, assays, and data-processing tools change, and as HuBMAP interacts and collaborates with other

related efforts, as they have for other consortium projects[26–31]. Where possible, metadata related to preanalytical variables (for example, annotations and nomenclature) and technologies will be harmonized, and protocols and standards will be shared with the wider research community.

## Building an integrated tissue map across scales

The diversity of data generated by HuBMAP, ranging across macroscopic and microscopic scales (for example, anatomical, histological, cellular, molecular and genomic) and multiple individuals, is essential to its core mission. Exploring each of these valuable datasets collectively will yield an integrated view of the human body. Hence, HuBMAP will develop analytical and visualization tools to bridge spatial and molecular relationships in order to help to generate a high-resolution three-dimensional molecular atlas of the human body.

The volume of data generated and collected by HuBMAP will require the utilization, extension and development of tools and pipelines for data processing. While we expect that initial data-processing tools will be based on methods developed by consortium members, HuBMAP will also work with and incorporate algorithms developed by other programs and the wider research community to supplement, enhance or update its pipelines. To this end, HuBMAP will develop one or more portals tailored to emerging use cases identified through a series of user needs. These open source portals will use recognized standards and be interoperable with other platforms, such as the HCA Data Coordination Platform, making it possible to readily add, update, and use new software modules (for example, as with Dockstore[32] and Toil[33]). The portion of HuBMAP data that will be open source can live on or be accessed from multiple platforms, enhancing its utility. This infrastructure will enable external developers to apply their codes, applications, open application programming interfaces, and data schema to facilitate customized processing and analysis of HuBMAP data in concert with other data sources. Furthermore, by actively working with other global and NIH initiatives, the consortium will seek to
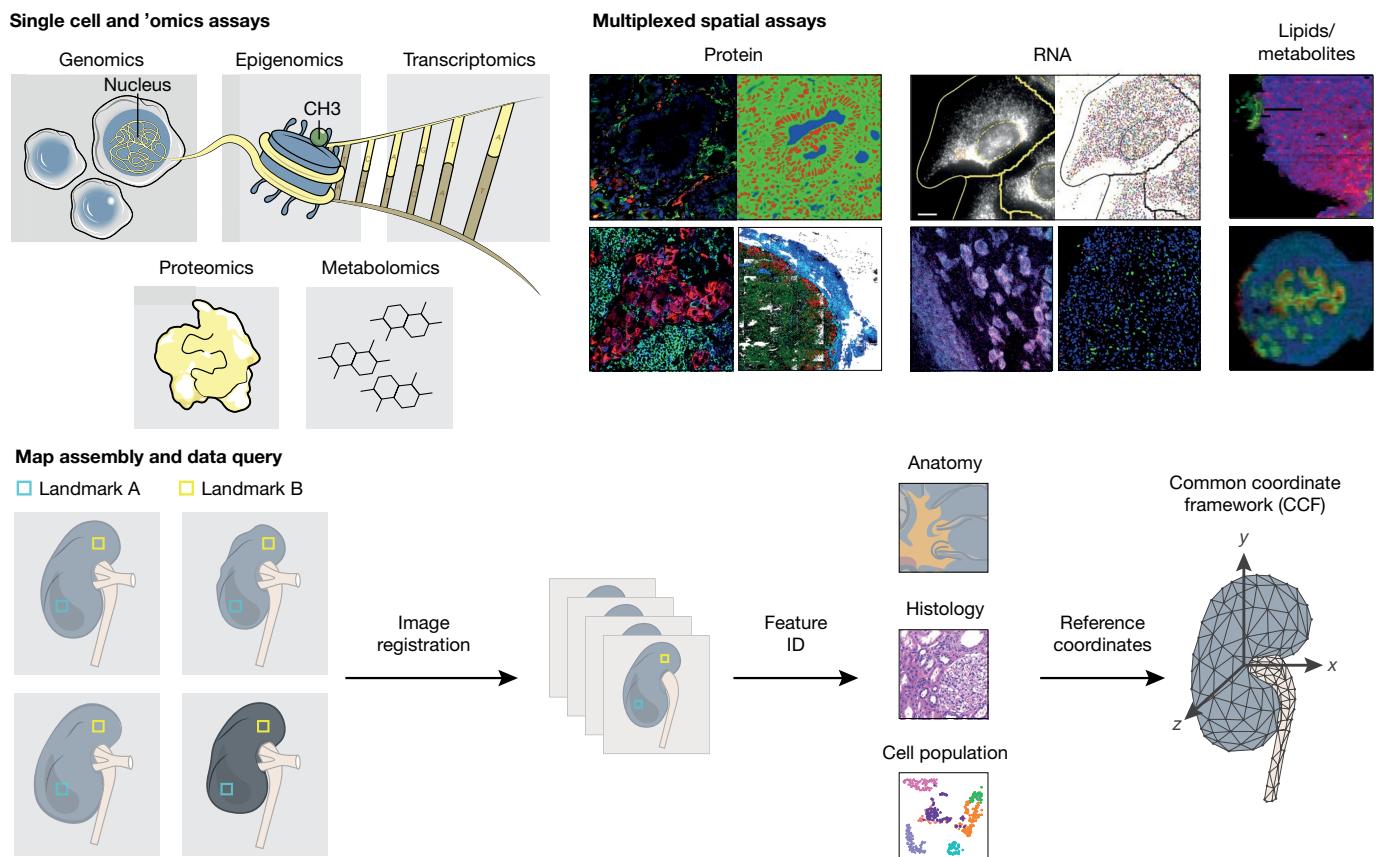
**Fig. 3 | Map generation and assembly across cellular and spatial scales.** HuBMAP aims to produce an atlas in which users can refer to a histological slide from a specific part of an organ and, in any given cell, understand its contents on multiple 'omic levels—genomic, epigenomic, transcriptomic, proteomic, and/or metabolomic. To achieve these ends, centres will apply a combination of imaging, 'omics and mass spectrometry techniques to specimens collected in a reproducible manner from specific sites in the body. These data will be then be integrated to arrive at a high-resolution, high-content three-dimensional map for any given tissue. To ensure inter-individual differences will not be confounded with collection heterogeneity, a robust CCF will be developed.

reduce the barriers to browsing, searching, aggregating, and analysing data across portals and platforms.

To fully integrate spatial and molecular data across individuals, HuBMAP will create a common coordinate framework (CCF) that defines a three-dimensional spatial representation, leveraging both an early consortium-wide effort to standardize technologies and assays using a single common tissue and the broader range of tissues of the human body analysed across multiple scales (whole body to single cells). This spatial representation will serve as an addressable scaffold for all HuBMAP data, enabling unified interactive exploration and visualization (search, filter, details on demand) and facilitating comparative analysis across individuals, technologies, and laboratories[34,35]. To achieve these objectives, HuBMAP envisions a strategy inspired by other tissue atlas efforts[36–38] that leverages the identification of 'landmark' features, including key anatomical structures and canonical components of tissue organization (for example, epidermal boundaries and normally spatially invariant vasculature) that can be identified in all individuals. These landmarks will enable a 'semi-supervised' strategy for aligning and assembling an integrated reference, upon which HuBMAP investigators can impose diverse coordinate systems, including relative representations and zone-based projections. As one example, an open-source, computational histology topography cytometry analysis toolbox (histoCAT[39]) currently facilitates two-dimensional visualization and will soon also be applicable to three-dimensional reconstruction. Ontology-based frameworks will be explored in parallel to effectively categorize, navigate, and name multiscale data; synergies are expected between these two approaches. Whenever available, medical imaging, such as CT and MRI information, will serve as a basis for landmarking and constructing the CCF.

## Technology development and implementation

Quantitative imaging of different classes of biomolecule in the same tissue sample with high spatial resolution, sensitivity, specificity, and throughput is central to the development of detailed tissue maps. Although no single technique can fully address this challenge at present, the development and subsequent multiplexing of complementary capabilities provides a promising approach for accelerating tissue mapping efforts. The HuBMAP innovation technologies groups aim to develop several innovative approaches that will address the limitations of existing state-of-the-art techniques. For example, transformative technologies such as signal amplification by exchange reaction (SABER)[40,41], seqFISH[18,42,43], and Lumiphore probes[44] will be refined to improve multiplexing, sensitivity, and throughput for imaging RNA and proteins across multiple tissues. Furthermore, new mass spectrometry imaging techniques will enable the quantitative mapping of hundreds of lipids, metabolites, and proteins from the same tissue section with high spatial resolution and sensitivity[45,46]. There is also scope within the program to develop and test new technologies. These efforts will benefit from the development of new computational tools and machine learning algorithms, optimized first from data generated from a common tissue during the pilot phase, for data integration across modalities.

## Challenges

Previous programs such as GTEx[5] have faced the challenge of optimizing the collection, preservation, and processing of a wide variety of tissue types from multiple donors. However, one of the goals of HuBMAP, to generate comprehensive, interactive high-resolution maps using a wide variety of assays, introduces an added level of complexity. Mapping functionally important biomolecules, including some of

which we may not even be aware and for which sensitive, specific, and high-throughput assays are still lacking, will require close attention. Moreover, the program will produce an unprecedented volume and diversity of datasets for comprehensive data capture, management, mining, modelling, visual exploration and communication. The integration of data from different modalities is required for generating robust maps; it will be necessary to develop corresponding analysis and interactive visualization tools to ensure that the data and atlas are accessible to the entire life-sciences community. Finally, given the enormity of a human atlas, HuBMAP faces the challenges of prioritization of tissues and technologies, sampling across tissues and donors, and optimally synergizing its efforts with international efforts. Determining the number of cells, fields of view, and samples needed to capture rare cell types, states or tissue structures is an important challenge, but can be tackled with adaptive power analyses, leveraging the growing amount of data available both within HuBMAP and from other consortia as well as individual groups.

### Resources and community engagement

HuBMAP is an important part of the international mission to build a high-resolution cellular and spatial map of the human body, and we are firmly committed to close collaboration and synergy with the aforementioned initiatives to build an easy-to-use platform and interoperable datasets that will accelerate the realization of a high-resolution human atlas. Shared guiding principles around open data, tools, and access will enable collaborative and integrated analyses of data produced by diverse consortia. To achieve this synergy, HuBMAP and other consortia will work together to tackle common computational challenges, such as cellular annotation, through formal and informal gatherings focused on addressing these problems, planned joint benchmarking and hands-on jamborees and workshops. Another example of the potential for close collaboration is in the study of the colon; multiple projects funded by HuBMAP, the Human Tumour Atlas Network, and the Wellcome Trust will be complemented by projects funded by the Leona M. and Harry B. Helmsley Charitable Trust. With projects focusing on partly distinct regions and diseases (for example, normal tissue, colon cancer, and Crohn's disease), it will be important for all of the programs to ensure that data are collected and made available in a consistent manner, and HuBMAP will play an active part in such efforts. As a concrete next step, HuBMAP, in collaboration with other NIH programs, plans to hold a joint meeting with the Human Cell Atlas initiative to identify and work on areas of harmonization and collaboration during the spring of 2020. In parallel, HuBMAP participants engage in the meetings and activities of other consortia, such as the Human Cell Atlas or the Human Tumour Atlas Network, thus forming tight connections. We have started a series of open meetings to develop the CCF, with the first of these recently held in collaboration with the Kidney Precision Medicine Program and focused on the kidney.

HuBMAP will provide capabilities for data submission, access, and analysis following FAIR (findable, accessible, interoperable, and reusable) data principles[47]. We will develop policies for prompt and regular data releases in commonly used formats, consistent with similar initiatives. We anticipate that the first round of data will be released in the summer of 2020, with subsequent releases at timely intervals thereafter. Robust metadata will comprise all aspects of labelling and provenance, including de-identified donor information (both demographic and clinical), details of tissue processing and protocols, data levels, and processing pipelines.

Indeed, engagement and outreach to the broader scientific community and other mapping centres is central to ensure that resources generated by HuBMAP will be leveraged broadly for sustained impact. To ensure that browsers and visualization tools from HuBMAP are valuable, the consortium will work closely with anatomists, pathologists, and visualization and user experience experts, including those with expertise in virtual or augmented reality. As described above, we expect that the diversity of normal samples included in this project will facilitate valuable comparative analyses, pinpointing how cells and tissue structures vary across individuals, throughout the lifespan, and in the emergence of dysfunction and disease. The program will build its resources with these use cases in mind and provide future opportunities, such as the demonstration projects, for close collaboration with domain experts. We also anticipate that these data will be highly useful for the generation of new biomedical hypotheses, tissue engineering, the development of robust simulations of spatiotemporal interactions, machine learning of tissue features, and educational purposes.

### Conclusions

Analogous to the release of the first human genome build, we anticipate that the first reference three-dimensional tissue maps will represent the tip of the iceberg in terms of their ultimate scope and eventual impact. HuBMAP, working closely with other initiatives, aspires to help to build a foundation by generating a high-resolution atlas of key organs in the normal human body and capturing inter-individual differences, as well as acting as a key resource for new contributions in the growing fields of tissue biology and cellular ecosystems. Given the focus of HuBMAP on spatial molecular mapping, the consortium will contribute to the community of efforts seeking similar goals, with a special emphasis on providing leadership in the development of analytical methods for its data types and for developing a common coordinate framework to integrate data. Ultimately, we hope to catalyse novel views on the organization of tissues, regarding not only which types of cells are neighbouring one another, but also the gene and protein expression patterns that define these cells, their phenotypes, and functional interactions. In addition to encouraging the establishment of intra- and extra-consortium collaborations that align with HuBMAP's overall mission, we envision an easily accessible, publicly available user interface through which data can be used to visualize molecular landscapes at the single-cell level, pathways and networks for molecules of interest, and spatial and temporal changes across a given cell type of interest. Researchers will also be able to browse, search, download, and analyse the data in standard formats with rich metadata that, over time, will enable users to query and analyse datasets across similar programs.

Importantly, we believe that the project's compilation of different types of multi-omic information at the single-cell level in a spatially resolved manner will represent an important step in the advancement of our understanding of human biology and precision medicine. These data have the potential to redefine types or subtypes of cells and their relationships within and between tissues beyond the traditional understanding that can be obtained through standard methods (for example, microscopy and flow cytometry). We hope this work will be part of a foundation that enables diagnostic interrogation, modelling, navigation, and targeted therapeutic interventions at such an unprecedented resolution to be transformative for the biomedical field.

1. Svensson, V. et al. Power analysis of single-cell RNA-sequencing experiments. *Nat. Methods* **14**, 381–387 (2017).
2. Schwartzman, O. & Tanay, A. Single-cell epigenomics: techniques and emerging applications. *Nat. Rev. Genet.* **16**, 716–726 (2015).
3. Tanay, A. & Regev, A. Scaling single-cell genomics from phenomenology to mechanism. *Nature* **541**, 331–338 (2017).
4. Norris, J. L. & Caprioli, R. M. Analysis of tissue specimens by matrix-assisted laser desorption/ionization imaging mass spectrometry in biological and clinical research. *Chem. Rev.* **113**, 2309–2342 (2013).
5. GTEx Consortium. Genetic effects on gene expression across human tissues. *Nature* **550**, 204–213 (2017).
6. National Research Council of the National Academies. *Enhancing the Effectiveness of Team Science* (National Academies Press, 2015).
7. Regev, A. et al. The Human Cell Atlas. *eLife* **6**, e27041 (2017).
8. Interactive human protein atlas launches. *Cancer Discov.* **5**, 339 (2015).
9. Ecker, J. R. et al. The BRAIN Initiative Cell Census Consortium: lessons learned toward generating a comprehensive brain cell atlas. *Neuron* **96**, 542–557 (2017).
10. NCI Board of Scientific Advisors and the National Cancer Advisory Board. Human Tumor Atlas (HTA) Network. https://www.cancer.gov/research/key-initiatives/moonshot-cancer-initiative/funding/upcoming/hta-foa-video (National Cancer Institute, 2017).

11. Cao, J. et al. Comprehensive single-cell transcriptional profiling of a multicellular organism. *Science* **357**, 661–667 (2017).
12. Buenrostro, J. D. et al. Single-cell chromatin accessibility reveals principles of regulatory variation. *Nature* **523**, 486–490 (2015).
13. Cusanovich, D. A. et al. Multiplex single cell profiling of chromatin accessibility by combinatorial cellular indexing. *Science* **348**, 910–914 (2015).
14. Cusanovich, D. A. et al. A single-cell atlas of in vivo mammalian chromatin accessibility. *Cell* **174**, 1309–1324 (2018).
15. Shah, S., Lubeck, E., Zhou, W. & Cai, L. In situ transcription profiling of single cells reveals spatial organization of cells in the mouse hippocampus. *Neuron* **92**, 342–357 (2016).
16. Goltsev, Y. et al. Deep profiling of mouse splenic architecture with CODEX multiplexed imaging. *Cell* **174**, 968–981 (2018).
17. Lubeck, E., Coskun, A. F., Zhiyentayev, T., Ahmad, M. & Cai, L. Single-cell in situ RNA profiling by sequential hybridization. *Nat. Methods* **11**, 360–361 (2014).
18. Eng, C. L. et al. Transcriptome-scale super-resolved imaging in tissues by RNA seqFISH. *Nature* **568**, 235–239 (2019).
19. Van de Plas, R., Yang, J., Spraggins, J. & Caprioli, R. M. Image fusion of mass spectrometry and microscopy: a multimodality paradigm for molecular tissue mapping. *Nat. Methods* **12**, 366–372 (2015).
20. Spraggins, J. M. et al. Next-generation technologies for spatial proteomics: Integrating ultra-high speed MALDI-TOF and high mass resolution MALDI FTICR imaging mass spectrometry for protein analysis. *Proteomics* **16**, 1678–1689 (2016).
21. Rapsomaniki, M. A. et al. CellCycleTRACER accounts for cell cycle and volume in mass cytometry data. *Nat. Commun.* **9**, 632 (2018).
22. Chevrier, S. et al. Compensation of signal spillover in suspension and imaging mass cytometry. *Cell Syst.* **6**, 612–620 (2018).
23. Schulz, D. et al. Simultaneous multiplexed imaging of mRNA and proteins with subcellular resolution in breast cancer tissue samples by mass cytometry. *Cell Syst.* **6**, 25–36 (2018).
24. Chevrier, S. et al. An immune atlas of clear cell renal cell carcinoma. *Cell* **169**, 736–749 (2017).
25. Patterson, N. H. et al. Next generation histology-directed imaging mass spectrometry driven by autofluorescence microscopy. *Anal. Chem.* **90**, 12404–12413 (2018).
26. Teng, M. et al. A benchmark for RNA-seq quantification pipelines. *Genome Biol.* **17**, 74 (2016).
27. Petryszak, R. et al. Expression Atlas update—an integrated database of gene and protein expression in humans, animals and plants. *Nucleic Acids Res.* **44**, D746–D752 (2016).
28. Hong, E. L. et al. Principles of metadata organization at the ENCODE data coordination center. *Database (Oxford)* **2016**, baw001 (2016).
29. Campbell-Thompson, M. et al. Network for Pancreatic Organ Donors with Diabetes (nPOD): developing a tissue biobank for type 1 diabetes. *Diabetes Metab. Res. Rev.* **28**, 608–617 (2012).
30. Pugliese, A. et al. The Juvenile Diabetes Research Foundation Network for Pancreatic Organ Donors with Diabetes (nPOD) Program: goals, operational model and emerging findings. *Pediatr. Diabetes* **15**, 1–9 (2014).
31. Philips, T. et al. Factors that influence the quality of RNA from the pancreas of organ donors. *Pancreas* **46**, 252–259 (2017).
32. O'Connor, B. D. et al. The Dockstore: enabling modular, community-focused sharing of Docker-based genomics tools and workflows. *F1000Res.* **6**, 52 (2017).
33. Vivian, J. et al. Toil enables reproducible, open source, big biomedical data analyses. *Nat. Biotechnol.* **35**, 314–316 (2017).
34. Butler, A., Hoffman, P., Smibert, P., Papalexi, E. & Satija, R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat. Biotechnol.* **36**, 411–420 (2018).
35. Haghverdi, L., Lun, A. T. L., Morgan, M. D. & Marioni, J. C. Batch effects in single-cell RNA-sequencing data are corrected by matching mutual nearest neighbors. *Nat. Biotechnol.* **36**, 421–427 (2018).
36. Mazziotta, J. et al. A probabilistic atlas and reference system for the human brain: International Consortium for Brain Mapping (ICBM). *Phil. Trans. R. Soc. Lond. B* **356**, 1293–1322 (2001).
37. Fonseca, C. G. et al. The Cardiac Atlas Project—an imaging database for computational modeling and statistical atlases of the heart. *Bioinformatics* **27**, 2288–2295 (2011).
38. Fürth, D. et al. An interactive framework for whole-brain maps at cellular resolution. *Nat. Neurosci.* **21**, 139–149 (2018).
39. Schapiro, D. et al. histoCAT: analysis of cell phenotypes and interactions in multiplex image cytometry data. *Nat. Methods* **14**, 873–876 (2017).
40. Saka, S. K. et al. Immuno-SABER enables highly multiplexed and amplified protein imaging in tissues. *Nat. Biotechnol.* **37**, 1080–1090 (2019).
41. Kishi, J. Y. et al. SABER amplifies FISH: enhanced multiplexed imaging of RNA and DNA in cells and tissues. *Nat. Methods* **16**, 533–544 (2019).
42. Shah, S., Lubeck, E., Zhou, W. & Cai, L. seqFISH accurately detects transcripts in single cells and reveals robust spatial organization in the hippocampus. *Neuron* **94**, 752–758 (2017).
43. Zhu, Q., Shah, S., Dries, R., Cai, L. & Yuan, G.-C. Identification of spatially associated subpopulations by combining scRNAseq and sequential fluorescence in situ hybridization data. *Nat. Biotechnol.* **36**, 1183–1190 (2018).
44. Cho, U. et al. Ultrasensitive optical imaging with lanthanide lumiphores. *Nat. Chem. Biol.* **14**, 15–21 (2018).
45. Yin, R. et al. High spatial resolution imaging of mouse pancreatic islets using nanospray desorption electrospray ionization mass spectrometry. *Anal. Chem.* **90**, 6548–6555 (2018).
46. Zhu, Y. et al. Nanodroplet processing platform for deep and quantitative proteome profiling of 10–100 mammalian cells. *Nat. Commun.* **9**, 882 (2018).
47. Wilkinson, M. D. et al. The FAIR Guiding Principles for scientific data management and stewardship. *Sci. Data* **3**, 160018 (2016).

**Author contributions** M.P.S., S.L., A.L.P., M.A.A, J.R., O.R.-R., A.H., R.S., N.G., J.S., J.L., P.H., N.A.N., J.C.S., Z.B.-J., K.Z., K.B., Y.L., R.C., D.P., A.L.R., A.P., M.B., and Z.S.G. wrote the manuscript; A.R. and L.G. generated the figures and reviewed and/or edited the manuscript; J.R., O.R.-R., A.H., R.S., N.G., J.S., J.L., P.H., N.A.N., J.C.S., Z.B.-J., K.Z., K.B., Y.L., R.C., D.P., A.L.R., A.P., M.B., and Z.S.G. contributed to discussion and provided critical review and/or revision of the manuscript. All other co-authors outside the writing group reviewed the manuscript and approved of its submission for publication. Contacts of principal investigators for respective TMCs, TTDs, or HIVE are listed first in each section.

**Writing Group** Michael P. Snyder[1]*, Shin Lin[2]*, Amanda Posgai[3]*, Mark Atkinson[3]*, Aviv Regev[4,5], Jennifer Rood[4], Orit Rozenblatt-Rosen[4], Leslie Gaffney[4], Anna Hupalowska[4], Rahul Satija[6,7]*, Nils Gehlenborg[8]*, Jay Shendure[9], Julia Laskin[10]*, Pehr Harbury[11]*, Nicholas A. Nystrom[12]*, Jonathan C. Silverstein[13]*, Ziv Bar-Joseph[14]*, Kun Zhang[15]*, Katy Börner[16]*, Yiing Lin[17], Richard Conroy[18]*, Dena Procaccini[18], Ananda L. Roy[18], Ajay Pillai[19], Marishka Brown[20] & Zorina S. Galis[20]

**Caltech-UW TMC** Long Cai[21]*, Jay Shendure[9], Cole Trapnell[9], Shin Lin[2]* & Dana Jackson[9]

**Stanford-WashU TMC** Michael P. Snyder[1]*, Garry Nolan[22], William James Greenleaf[1], Yiing Lin[17], Sylvia Plevritis[23], Sara Ahadi[1], Stephanie A. Nevins[1], Hayan Lee[1], Christian Martijn Schuerch[22], Sarah Black[22], Vishal Gautham Venkataraaman[22], Ed Esplin[1], Aaron Horning[1] & Amir Bahmani[1]

**UCSD TMC** Kun Zhang[15]*, Xin Sun[15], Sanjay Jain[24], James Hagood[25], Gloria Pryhuber[26] & Peter Kharchenko[8]

**University of Florida TMC** Mark Atkinson[3]*, Bernd Bodenmiller[27], Todd Brusko[3], Michael Clare-Salzler[3], Harry Nick[28], Kevin Otto[29], Amanda Posgai[3]*, Clive Wasserfall[3], Marda Jorgensen[3], Maigan Brusko[3] & Sergio Maffioletti[27]

**Vanderbilt University TMC** Richard M. Caprioli[30]*, Jeffrey M. Spraggins[30]*, Danielle Gutierrez[30], Nathan Heath Patterson[30], Elizabeth K. Neumann[30], Raymond Harris[31], Mark deCaestecker[31], Agnes B. Fogo[32], Raf van de Plas[33] & Ken Lau[34]

**California Institute of Technology TTD** Long Cai[21]*, Guo-Cheng Yuan[35], Qian Zhu[35] & Ruben Dries[35]

**Harvard TTD** Peng Yin[36,37]*, Sinem K. Saka[36,37], Jocelyn Y. Kishi[36,37], Yu Wang[36,37] & Isabel Goldaracena[36,37]

**Purdue TTD** Julia Laskin[10]*, DongHye Ye[10,38], Kristin E. Burnum-Johnson[39], Paul D. Piehowski[39], Charles Ansong[39] & Ying Zhu[39]

**Stanford TTD** Pehr Harbury[11]*, Tushar Desai[40], Jay Mulye[11], Peter Chou[11] & Monica Nagendran[40]

**HuBMAP Integration, Visualization, and Engagement (HIVE) Collaboratory:**
**Carnegie Mellon, Tools Component** Ziv Bar-Joseph[14]*, Sarah A. Teichmann[41]*, Benedict Paten[42]*, Robert F. Murphy[14], Jian Ma[14], Vladimir Yu. Kiselev[41], Carl Kingsford[14], Allyson Ricarte[14], Maria Keays[41], Sushma A. Akoju[14] & Matthew Ruffalo[14]

**Harvard Medical School, Tools Component** Nils Gehlenborg[8]*, Peter Kharchenko[8], Margaret Vella[8] & Chuck McCallum[8]

**Indiana University Bloomington, Mapping Component** Katy Börner[16]*, Leonard E. Cross[16], Samuel H. Friedman[43], Randy Heiland[16], Bruce Herr II[16], Paul Macklin[16], Ellen M. Quardokus[16], Lisel Record[16], James P. Sluka[16] & Griffin M. Weber[8]

**Pittsburgh Supercomputing Center and University of Pittsburgh, Infrastructure and Engagement Component** Nicholas A. Nystrom[12]*, Jonathan C. Silverstein[13]*, Philip D. Blood[12], Alexander J. Ropelewski[12], William E. Shirey[13] & Robin M. Scibek[12]

**University of South Dakota, Collaboration Core** Paula Mabee[44]*, W. Christopher Lenhardt[45], Kimberly Robasky[45,46,47] & Stavros Michailidis[48]

**New York Genome Center, Mapping Component** Rahul Satija[6,7]*, John Marioni[41,49], Aviv Regev[4,5], Andrew Butler[6,7], Tim Stuart[6], Eyal Fisher[50], Shila Ghazanfar[50], Jennifer Rood[4], Leslie Gaffney[4], Gokcen Eraslan[4], Tommaso Biancalani[4] & Eeshit D. Vaishnav[4]

**NIH HuBMAP Working Group** Richard Conroy[18]*, Dena Procaccini[18], Ananda Roy[18], Ajay Pillai[19], Marishka Brown[20], Zorina Galis[20], Pothur Srinivas[20], Aaron Pawlyk[51], Salvatore Sechi[51], Elizabeth Wilder[18] & James Anderson[18]

[1]Department of Genetics, Stanford School of Medicine, Stanford, CA, USA. [2]Department of Medicine, University of Washington, Seattle, WA, USA. [3]Department of Pathology, University of Florida Diabetes Institute, Gainesville, FL, USA. [4]Klarman Cell Observatory Broad Institute of MIT and Harvard, Cambridge, MA, USA. [5]Howard Hughes Medical Institute, Koch Institute of Integrative Cancer Research, Department of Biology, Massachusetts Institute of Technology, Cambridge, MA, USA. [6]New York Genome Center, New York, NY, USA. [7]New York University, New York, NY, USA. [8]Department of Biomedical Informatics, Harvard Medical School, Boston, MA, 02115, USA. [9]Brotman Baty Institute for Precision Medicine, Allen Discovery Center for Cell Lineage Tracing, Howard Hughes Medical Institute, Department of Genome Sciences, University of Washington, Seattle, WA, USA. [10]Department of Chemistry, Purdue University, West Lafayette, IN, USA. [11]Department of Biochemistry, Stanford University School of Medicine, Stanford, CA, USA. [12]Pittsburgh Supercomputing Center, Carnegie Mellon University, Pittsburgh, PA, USA. [13]Department of Biomedical Informatics, University of Pittsburgh, Pittsburgh, PA, USA. [14]Computational Biology Department, School of Computer Science, Carnegie Mellon University, Pittsburgh, PA, USA. [15]Department of Bioengineering, University of California San Diego, La Jolla, CA, USA. [16]Department of Intelligent Systems Engineering, School of Informatics, Computing, and Engineering, Indiana University, Bloomington, IN, USA. [17]Department of Surgery, Washington University School of Medicine, St Louis, MO, USA. [18]Office of Strategic Coordination, Division of Program Coordination, Planning, and Strategic Initiatives, National Institutes of Health, Bethesda, MD, USA. [19]National Human Genome Research Institute, National Institutes of Health, Bethesda, MD, USA. [20]National Heart, Lung, and Blood Institute, National Institutes of Health, Bethesda, MD, USA. [21]Department of Biology and Biological Engineering, California Institute of Technology, Pasadena, CA, USA. [22]Department of Microbiology, Stanford School of Medicine, Stanford, CA, USA. [23]Department of Radiology, Stanford School of Medicine, Stanford, CA, USA. [24]Department of Medicine, Washington University in St Louis, St Louis, MO, USA. [25]Department of Pediatrics, University of North Carolina School of Medicine, Chapel Hill, NC, USA. [26]Department of Pediatrics, University of Rochester, Rochester, NY, USA. [27]Institute of Molecular Life Sciences, University of Zurich, Zurich, Switzerland. [28]Department of Neuroscience, University of Florida, Gainesville, FL, USA. [29]Department of Biomedical Engineering, University of Florida, Gainesville, FL, USA. [30]Mass Spectrometry Research Center, Department of Biochemistry, Vanderbilt University, Nashville, TN, USA. [31]Department of Medicine, Vanderbilt University Medical Center, Nashville, TN, USA. [32]Department of Pathology, Microbiology and Immunology, Vanderbilt University Medical Center, Nashville, TN, USA. [33]Delft Center for Systems and Control, Delft University of Technology, Delft, The Netherlands. [34]Department of Cell and Developmental Biology, Vanderbilt University, Nashville, TN, USA. [35]Department of Biostatistics and Computational Biology, Dana-Farber Cancer Institute, Boston, MA, USA. [36]Wyss Institute for Biologically Inspired Engineering, Harvard University, Boston, MA, USA. [37]Department of Systems Biology, Harvard Medical School, Boston, MA, USA. [38]Department of Electrical and Computer Engineering, Opus College of Engineering, Marquette University, Milwaukee, WI, USA. [39]Biological Sciences Division, Pacific Northwest National Laboratory, Richland, WA, USA. [40]Department of Internal Medicine, Division of Pulmonary & Critical Care, Stanford University School of Medicine, Stanford, CA, USA. [41]Cellular Genetics Programme, Wellcome Sanger Institute, Hinxton, UK. [42]Department of Biomolecular Engineering, Jack Baskin School of Engineering, University of California Santa Cruz, Santa Cruz, CA, USA. [43]Opto-Knowledge Systems, Torrance, CA, USA. [44]Department of Biology, University of South Dakota, Vermillion, SD, USA. [45]Renaissance Computing Institute, University of North Carolina, Chapel Hill, NC, USA. [46]Department of Genetics, University of North Carolina, Chapel Hill, NC, USA. [47]School of Information and Library Science, University of North Carolina, Chapel Hill, NC, USA. [48]Knowinnovation, Buffalo, NY, USA. [49]European Molecular Biology Laboratory, European Bioinformatics Institute (EMBL-EBI), Wellcome Genome Campus, Hinxton, UK. [50]Cancer Research UK Cambridge Institute, University of Cambridge, Cambridge, UK. [51]National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, MD, USA. *e-mail: mpsnyder@stanford.edu; shinlin@uw.edu; aposgai6@ufl.edu; atkinson@pathology.ufl.edu; rsatija@nygenome.org; nils@hms.harvard.edu; jlaskin@purdue.edu; harbury@stanford.edu; nystrom@psc.edu; j.c.s@pitt.edu; zivbj@cs.cmu.edu; kzhang@bioeng.ucsd.edu; katy@indiana.edu; conroyri@mail.nih.gov; lcai@caltech.edu; richard.m.caprioli@vanderbilt.edu; jeff.spraggins@vanderbilt.edu; peng_yin@hms.harvard.edu; st9@sanger.ac.uk; bpaten@ucsc.edu; paula.mabee@usd.edu

# Real-time discrimination of earthquake foreshocks and aftershocks

Laura Gulia[1]* & Stefan Wiemer[1]

Immediately after a large earthquake, the main question asked by the public and decision-makers is whether it was the mainshock or a foreshock to an even stronger event yet to come. So far, scientists can only offer empirical evidence from statistical compilations of past sequences, arguing that normally the aftershock sequence will decay gradually whereas the occurrence of a forthcoming larger event has a probability of a few per cent. Here we analyse the average size distribution of aftershocks of the recent Amatrice–Norcia and Kumamoto earthquake sequences, and we suggest that in many cases it may be possible to discriminate whether an ongoing sequence represents a decaying aftershock sequence or foreshocks to an upcoming large event. We propose a simple traffic light classification to assess in real time the level of concern about a subsequent larger event and test it against 58 sequences, achieving a classification accuracy of 95 per cent.

All crustal moderate-to-large earthquakes are followed by a decaying aftershock sequence that typically lasts for years. In some cases, this decaying sequence is interrupted by an even larger, and often more destructive, subsequent mainshock. One of the biggest unknowns in real-time seismic hazard assessment during an ongoing seismic sequence is whether the largest event—the mainshock—has already happened or is still to come. There are no scientific means yet to prospectively distinguish between 'classical' aftershock sequences and a sequence of potential foreshocks to an upcoming larger event[1,2]—the latter being typically the biggest concern of the local population and civil protection authorities. So far, the only answer that science can offer to this first-order question is a purely statistical one, based on compilations of empirical observations: the chance that after a moderate earthquake an even larger event will occur within five days and 10 km is typically 5%[3,4]. These numbers are at the core of existing systems for operational earthquake forecasting[5,6] using algorithms such as ETAS (epidemic type aftershock sequences)[7,8] or STEP (short-term earthquake probability)[9].

From the physics point of view, the probability of a subsequent larger mainshock must depend on the stress conditions set up by the previous events and the long-term tectonic stress conditions[10,11]. These conditions, as well as the location of potential faults, are typically unknown, and physics-based approaches employing Coulomb stress transfer have so far not been successful in forecasting upcoming mainshocks any better than statistical models[12], whereas their information gain is typically too low to warrant action[13,14]. There have been a number of attempts to identify foreshocks using waveform analysis or other precursory phenomena[15–17] but these have not yet resulted in improved earthquake forecasting abilities.

Here we use the fact that the time after a moderate or larger mainshock is the most data-rich period during the earthquake cycle, with thousands of aftershocks (or potential foreshocks) occurring within hours. These events allow observing spatial and temporal transients at resolutions 1,000–10,000 times higher than during normal conditions. Measuring changes in the stress caused by the mainshock is possible only indirectly and with somewhat low precision. The average size distribution of earthquakes—that is, the $b$ value of the Gutenberg–Richter law[18,19] ($\log N = a - bM$ where $N$ is the cumulative number of events above magnitude $M$, $a$ describes the productivity and $b$ the average

size distribution of the earthquakes)—is sensitive to differential stress; its inverse dependence on differential stress has been confirmed many times in both laboratory[20–22] and field[23] studies. Recently, analysis of a stack of 58 aftershock sequences from California, Japan, Italy and Alaska—31 of them with data of good enough quality and sufficient abundance for subsequent stacking—showed that the $b$ value of aftershock sequences generally increases after the mainshock by 20%[24]. This study also presented a Coulomb stress-based model explaining the observed transients and their dependence on magnitude, distance and faulting styles.

We propose that sequences diverting from the generally observed increased $b$ value after a mainshock are those of high concern, for which a subsequent larger event is likely to occur. Therefore, real-time monitoring of the $b$ value in aftershock sequences can be used for real-time discrimination between foreshocks and aftershocks, allowing us to use a posteriori awareness for a priori alerts. Evidence supporting our hypothesis comes from investigating time series of two recent sequences: the $M = 6.6$ Norcia and $M = 7.3$ Kumamoto sequences, which occurred in 2016 and were preceded by subsequently identified foreshocks reaching magnitude 6.

## Establishing transients in $b$ values

Computing reliable time series of the $b$ value in aftershock zones is especially difficult, mostly because the quality, consistency and completeness of the seismicity catalogue is typically strongly affected by changes in the recording seismic network and by limitations in detection[25]. Therefore, the first hours or even days of data after a magnitude 6 event usually need to be excluded from the analysis, which is only feasible in areas with very good network coverage and advanced seismic data analysis procedures. In addition, it is often challenging to establish the local pre-mainshock $b$ values because of the sparseness of seismicity outside sequences and limitations in recording homogeneity. The detailed analysis procedure that we follow to compute the change in $b$ before and after the mainshock is described in Methods. We select events within 3 km of the fault plane because these have been shown to be the most reactive to stress changes[24].

***The Amatrice–Norcia sequences.*** On 24 August 2016, an earthquake with moment magnitude $M_w = 6.2$ struck on Amatrice, central
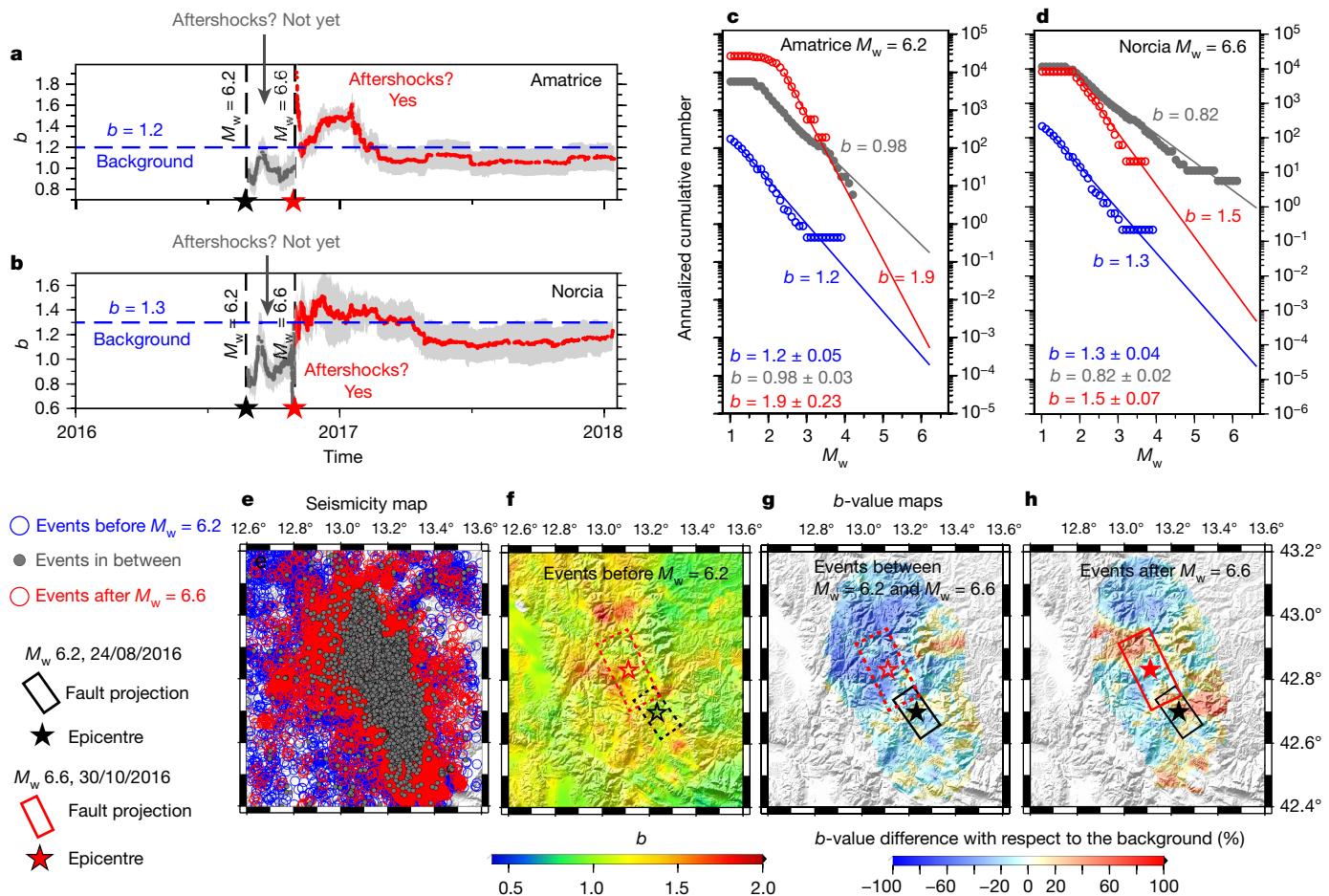
**Fig. 1 | Time–space analysis of b values for the Amatrice–Norcia sequence. a, b,** Time series of b values for the source regions of the Amatrice and Norcia mainshocks. The dashed blue lines indicate the background b values, and the vertical dashed grey lines represent the time of the $M_w = 6.2$ (Amatrice) and $M_w = 6.6$ (Norcia) earthquakes. The grey shaded areas show the uncertainty determined by bootstrapping (corresponding to one standard deviation). **c, d,** Frequency–magnitude distributions for the two source regions in three different periods (uncertainties from Shi and Bolt[47]). **e, f,** Seismicity maps, colour-coded by period (**e**) and pre-mainshock b value (**f**). **g, h,** Maps showing the change in the b value with respect to the background for the period between the two mainshocks (**g**) and the first two weeks of aftershocks (**h**).

Italy, killing about 300 people and severely damaging the town and neighbouring area. In contrast to most mainshocks in the region (for example, the 2009 $M_w = 6.3$ L'Aquila event[14]), the $M_w = 6.2$ event was not preceded by noticeable foreshocks. "Was this the main-shock?" is what the public, civil protection authorities and decision-makers were wondering at that time. It was not—two months later, on 30 October 2016, an $M_w = 6.6$ earthquake hit the town of Norcia, 20 km north-west of Amatrice, and neighbouring areas, revealing a posteriori that the $M_w = 6.2$ event and its 'aftershock' sequence were in fact foreshocks. This event was the strongest shock that occurred in the central–northern Apennines during the instru-mental era[26].

Assuming near-real-time conditions, we processed events from 2012 from the Italian earthquake catalogue that is homogeneous in terms of moment magnitudes[27]. We estimated a reference b value for the background (b = 1.2 for the interval between 2012 and the last event preceding the $M_w = 6.2$ earthquake). Using an automated quality and completeness analysis (see Methods), cross-checked by visual inspec-tion, we then removed the events in the first three days following the $M_w = 6.2$ event and computed the difference in b with respect to the background value (Fig. 1). After the 24 August 2016 $M_w = 6.2$ event, the b values near the Amatrice fault decreased by about 10%, from 1.2 to 1.1 (Fig. 1a)—a behaviour very different from the 20% increase observed generally. The plot of the frequency–size distribution of the earthquakes (Fig. 1c) shows that the decrease in b is stable according to high-quality data and does not depend on the chosen magnitude of completeness.

It also illustrates that the probability of a magnitude 6.6 event, inferred from the recurrence time[28], has increased by about a factor of 30. An even stronger decrease in b value is observed in the rupture area of the subsequent Norcia earthquakes (Fig. 1b), where the drop in b value is closer to 20% and the probability of a subsequent event of magni-tude 6.6 increases by a factor of 1,000 over the background (Fig. 1d). To analyse the spatial footprint of the change in b value, we map the percentage differences from the regional b value. We compute b values on a 2-km-spaced grid, sampling the nearest 250 events to each grid node and re-estimating the completeness in each node (see Methods). The mapping results are very consistent with the series analysis and frequency–magnitude distributions. In the time between the Amatrice and Norcia mainshocks, the b value decreased to the north of Amatrice by 20–50% (Fig. 1g).

The picture changes markedly after the $M_w = 6.6$ Norcia event; the b values in the Norcia and Amatrice source areas increase by 20–30% (Fig. 1a, b). Although the Norcia aftershock sequence includes many small events (owing to its larger magnitude), the chance of a subsequent larger event is substantially smaller than in the intervening period, close to the tectonic background rate. The differential map (Fig. 1h) also reveals that the b values increase in most regions. Analysis of the b values thus suggests that after the Norcia mainshock, typical aftershock activity is taking place, in agreement with the generic model. Indeed, until now (February 2019), no secondary large mainshock or larger event has taken place, although this was a highly concerning scenario in the autumn of 2016.
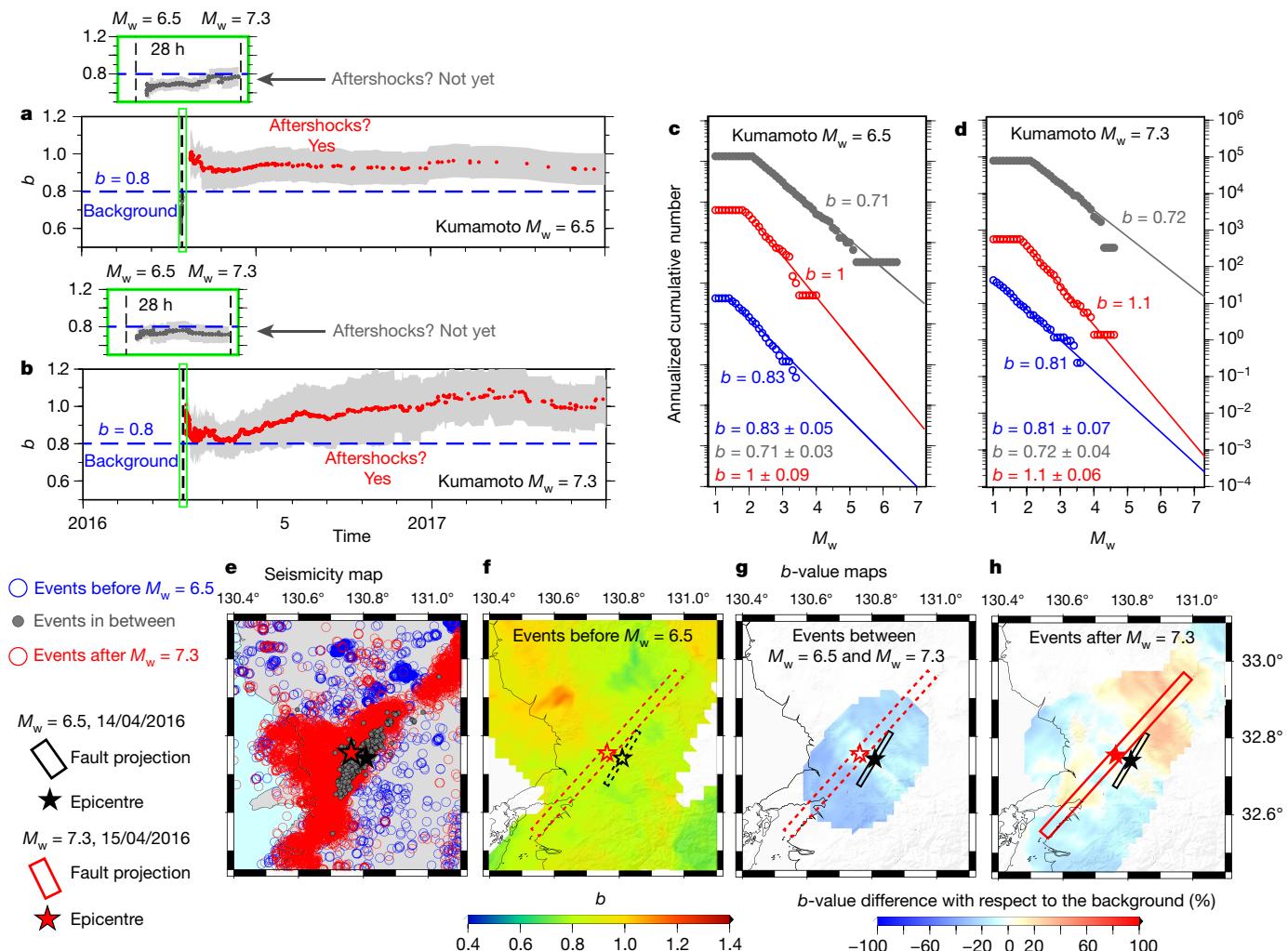
**Fig. 2 | Time–space analysis of *b* values for the Kumamoto sequence.**
**a, b**, Time series of *b* values for the source regions of the $M_w = 6.5$ and $M_w = 7.3$ events. The dashed blue line shows the background *b* values and the vertical dashed grey lines represent the time of the $M_w = 6.2$ and $M_w = 7.3$ earthquakes. The grey shaded areas represent the uncertainty determined by bootstrapping. **c, d**, Frequency magnitude distributions

for the two source regions in three different periods (uncertainty from Shi and Bolt[47]). **e, f**, Seismicity maps, colour-coded by period (**e**) and pre-mainshock *b* value (**f**). **g, h**, Map showing the change in the *b* value with respect to the background for the period between the two mainshocks (**g**) and the first two weeks of aftershocks (**h**).

*The Kumamoto sequence.* On 15 April 2016, an $M_w = 6.5$ earthquake occurred in the Kumamoto region, Japan[29], followed by a rich earthquake sequence considered to be aftershocks; 28 h later, an $M_w = 7.3$ earthquake revealed that these events were actually foreshocks. Both mainshocks caused severe damage. After the $M_w = 6.5$ earthquake, the Japan Meteorological Agency (JMA) warned of the possibility of large aftershocks with further damages. However, no information on an increased probability of $M_w = 7$ or larger earthquakes was made public because, according to the Earthquake Research Committee[30] protocol, the JMA had not considered the occurrence of larger earthquakes[31,32].

The Kumamoto sequence allows us to test our hypothesis in a different tectonic region and with mainshocks much closer together in time. We analyse the *b*-value time series for the source regions inferred for the $M_w = 6.5$ foreshock and for the $M_w = 7.3$ mainshock (Fig. 1a, b). For the background estimation, we select events in the JMA catalogue[29] starting in 2012 to avoid the first phase of the $M_w = 9$ Tohoku aftershocks. We divide the catalogue in three independent time periods: (1) from 2012 to the last event before the $M_w = 6.5$ earthquake (that is, the background), (2) from 1 h after the $M_w = 6.5$ earthquake to the last event before the mainshock and (3) from one day after the mainshock to the end of the catalogue (*b*-value time series) and the first two weeks of aftershocks (*b*-value map). The *b* values in the time

interval between the two shocks are similar or below the background level ($b = 0.7$), a result confirmed in the frequency–magnitude distributions (Fig. 2c) and in the differential map (Fig. 2g). Once the $M_w = 7.3$ event occurs, however, the *b* values of the subsequent events increase strongly by 20–40%. Consequently, whereas the annualized probability of an $M_w = 7.3$ event in the 28 h in period (2) increased by a factor of 1,000, it decreased after the second mainshock to almost background levels (Fig. 1d). Again, no subsequent large event has occurred so far.

**The 2011 Tohoku sequence**

The 2011 $M_w = 9$ Tohoku event and its $M_w = 7.3$ foreshock, recorded just two days before the mainshock, represent a further case study from a very different tectonic regime. The *b* values before and after the $M_w = 9$ earthquake have already been mapped by Tormann et al.[33]. The limits of the seismic network in precisely localizing off-shore events and the resulting scatter in hypocentres do not allow us to apply our method to the $M_w = 7.3$ box without increasing the selection radius to 12 km in order to have a sufficiently large dataset. We estimate a background *b* value of 0.62 (Fig. 3a), the *b* value in between (0.44; a 73% decrease) and the aftershock *b* value on the $M_w = 7.3$ fault (0.9; a 150% increase over the background). Differential *b*-value mapping cannot be performed owing to the paucity of events above the magnitude of completeness, $M_c$, in the short interval between the $M_w = 7.3$ and $M_w = 9$ events.
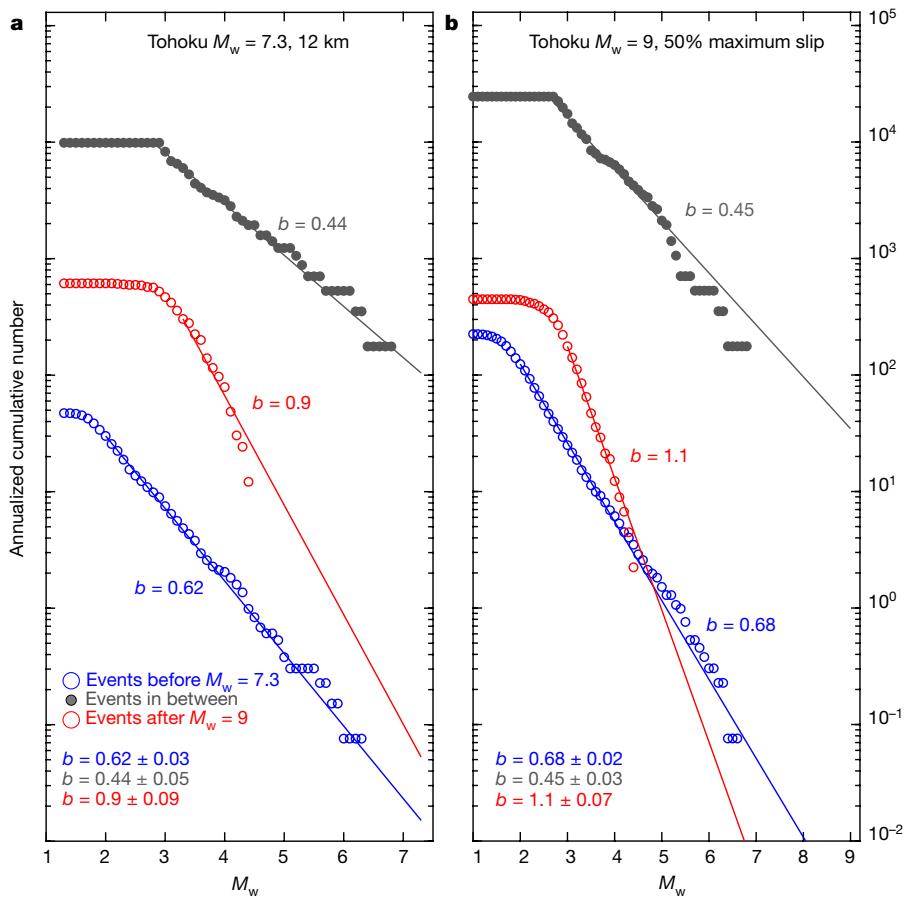
**Fig. 3 | Frequency–magnitude distributions for the Tohoku case study.** Frequency–magnitude distributions for three different periods of events within 12 km of the rupture plane of the $M_w = 7.3$ event (uncertainty from Shi and Bolt[47]). **b**, As in **a**, but for the events contained within 50% of the maximum slip contour of the $M_w = 9$ mainshock.

The $b$ value within 50% of the maximum slip contour behaved similarly (from 0.68 to 0.45 to 1.1; Fig. 3b).

**Possible physical mechanism**
Currently, there are two schools of thought regarding foreshock mechanisms and prognostic value[34,35]: (i) the deterministic point of view, which supports that foreshocks represent a precursory process, for example, a response to precursory slip on the fault[16,33,36], and (ii) the stochastic point of view, which considers foreshocks to be an indistinguishable part of earthquake clustering[17], described through a statistical process such as the ETAS model[7,8]. According to the ETAS model, there is no difference between foreshocks, mainshocks and aftershocks; all foreshocks are mainshocks with aftershocks that happen to be bigger. The rupture process is not cyclic but epidemic.

We interpret the observed drop in $b$ value documented in Figs. 1, 2 in a probabilistic framework of changes in the relevant stress conditions, which reconciles the aforementioned interpretations. Earthquakes of magnitude 6 and larger greatly perturb the stress field in the Earth crust. The amplitudes of the static- and dynamic-stress transfer decay with distance[37], and can both encourage and inhibit rupture. Under most conditions, this stress change will decrease the differential stress on nearby faults, thus increasing the $b$ value[24]. However, under certain conditions, the differential stress on nearby and already tectonically loaded faults can increase instead, leading to a drop in $b$ value and a subsequent much larger chance of an even stronger event. Conditions that favour such drops in $b$ value are probably the presence of critically pre-stressed faults and overall high stress levels, as well as a suitable orientation of the source and receiver faults. It is also possible that continued post-seismic slip, the impact of secondary aftershocks or precursory processes, such as deep precursory slip, may play a role.

Considering the numerous unknowns, we are currently unable to model individual sequences with sufficient reliability even a posteriori, and real-time modelling for warning purposes would be even more challenging. However, we can use the empirical observations of $b$-value changes as an input to improve earthquake risk mitigation.

**Towards real-time risk mitigation**
Our results (Figs. 1, 2) suggest that the evolution of $b$ values, analysed as a proxy for the average stress conditions of faults in the regions, can act as a first-order discriminator between normal aftershocks and likely precursory sequences. In the large majority of aftershock sequences, the $b$ value increases substantially after a mainshock of magnitude 6 or larger, typically by 20%[24]. This overall increase can be observed within hours of a mainshock, if indeed the seismic network is capable of reliable location and magnitude determination, and observing an increase in $b$ lowers the probability of a subsequent larger event by maybe an order of magnitude (Figs. 1d, 2d). If, on the other hand, $b$ remains the same or if it decreases considerably, then the probability of an even larger event is increased by several orders of magnitude.

We propose that our findings could be used to define a simple traffic light system expressing the level of concern associated with earthquakes. Traffic lights have been used to manage risk behaviour in a number of settings, such as food[38], health care[39], induced seismicity risk[40,41], volcanic eruption and in many other situations where decisions must be made in real time. They are a tool for recognizing risk in a quantitative way and then initiating risk reduction measures. The concept of our foreshock traffic light system (FTLS) is shown in Fig. 4. A yellow FTLS setting indicates that the $b$ value remains mostly unchanged in the aftershock sequence or is difficult to determine. We define yellow, somewhat arbitrarily, as a ±10% change from the
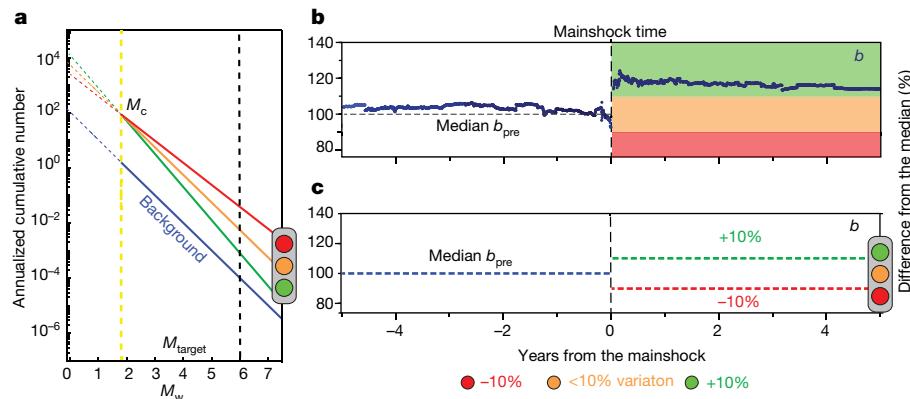
**Fig. 4 | The foreshock traffic light system. a, c**, Schematic representation of FTLS in the frequency–magnitude distribution view (**a**) and the $b$-value time series view (**c**). Green denotes an aftershock sequence with a $b$-value increase of about 20%, where no mainshock is expected; yellow indicates that the $b$ value remains unchanged in the aftershock sequence or is difficult to determine; red means that the $b$ value decreases considerably

and mitigation actions must be considered. $M_{target}$ is the reference value of $M$ in the example. **b**, $b$-value stacking of 31 sequences, showing the per cent difference with respect to the reference value (that is, the median of the background $b$ values, $b_{pre}$; black horizontal dashed line)[24]. The black vertical dashed line represents the time of the mainshock, shifted to zero.

background. Yellow represents the concern level according to present knowledge, with no additional discriminating information. A green FTLS setting corresponds to an aftershock sequence with an increase in $b$ value of 10% or more, and we postulate that more than 80% of all sequences will fall into this category. The ability to declare a green status would represent an important contribution to earthquake resilience, because it would greatly reduce uncertainty and concern and would allow a quicker return to normality, for example, by initiating rebuilding efforts. Finally, a red FTLS setting would be declared if the $b$ value decreased substantially, by 10% or more. In such situations, emergency managers should be especially concerned and consider actions such as continuing evacuations. In the future, it may be possible to refine the thresholds that we propose here on the basis of additional data, risk–cost–benefit analysis and by considering the uncertainty in $b$ values. We also suggest that spatially mapping relative changes in $b$ value may help us to define the most likely area of a subsequent large event. In the case of the Norcia event (Fig. 1), we note that the event did occur to the north of Amatrice, in the areas of the strongest decrease of $b$.

We tested our FTLS retrospectively on 58 sequences[24]. We calculated the percentage difference between the background and the $b$ value of the aftershocks, selecting all events within 3 km from the rupture area. This allowed us to obtain a robust value for 25 sequences, in addition to the values obtained for the foreshocks of the Norcia and Kumamoto sequences, resulting in a total of 29 sequences. Of these, 18 were classified as a green alert, 8 as yellow and 3 as red (Fig. 5). Only two of these mainshocks were followed by subsequent larger ones (Amatrice–Norcia and Kumamoto), which is in line with the 5% probability of a secondary

larger event[3,4]. We also added the values of the 2011 $M_w = 9$ Tohoku sequence (red alert after the $M_w = 7.3$ event, green after the $M_w = 9$ earthquake); however, we did not use them in the statistical analysis because, as discussed above, the method had been adjusted for larger hypocentre uncertainties.

For a first-order assessment of the performance of FTLS, we count in a binary classifier system the successful alerts (true positives), false alerts (false positives), missed events (false negatives) and correct negatives (true negatives). We consider yellow alerts as neutral. In this metric, we score two successful alerts, one false alert, no missed events and 18 correct negatives. Using confusion matrix analysis, we compute an accuracy of 0.95. If we assume that the chance of a subsequent larger event is 5%, then the random chance of correctly identifying two out of two mainshocks, with only one false alert and no missed events is below 1%.

The one false-positive red alert follows the $M_w = 6.2$ event in Morgan Hill (1984). Because it occurred 35 years ago, we speculate that the data quality may be inferior. The performance of the forecasting method can be further improved if we also analyse the spatial footprint of the $b$-value changes, as indicated in Figs. 1g, h, 2g, h. We perform FTLS classification also for the Norcia and Kumamoto ($M_w = 7.3$) source regions (which are known only a posteriori) of these mainshocks. Before the mainshock, $b$ decreases most sharply, by 55% in Norcia and 68% in Kumamoto ($M_w = 7.3$), resulting in two red alerts.

Our results demonstrate that changes in $b$ can act as a discriminant. Our hypothesis is also consistent with a physical framework in which stress influences the relative size distribution, and hence the probability,
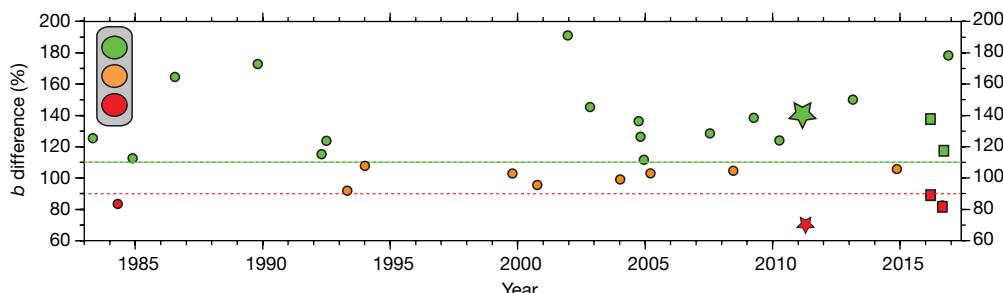


**Fig. 5 | Performance analysis of the proposed foreshock light system.** Difference between the $b$ values of aftershocks for 27 sequences recorded in California, Japan, Alaska and Italy (circles) and the background. Events are colour-coded by their FTLS class. The values obtained for the Amatrice and Kumamoto ($M_w = 6.5$) source regions (in between mainshocks and after the second mainshock) are shown as

red squares; green squares indicate the values obtained for Norcia and Kumamoto ($M_w = 7.3$). The plotted events are listed in Table 1. The green star shows the extrapolated value for Tohoku, 2011, from figure 3 in Tormann et al.[33], and the red star represents the $M_w = 7.3$ Tohoku foreshock shown in Fig. 3.

Table 1 | b-value differences δb for the sequences in Fig. 5.

| Name | Catalogue | Date | Mms | δb | FTLS colour |
|---|---|---|---|---|---|
| Coalinga | ANSS | 2 May 1983 | 6.7 | 125% | Green |
| Morgan Hill | ANSS | 24 Apr 1984 | 6.2 | 83% | Red |
| Round Valley | ANSS | 23 Nov 1984 | 6.1 | 113% | Green |
| Chalfant Valley | ANSS | 21 Jul 1986 | 6.4 | 164% | Green |
| Loma Prieta | ANSS | 18 Oct 1989 | 7 | 173% | Green |
| Joshua Tree | ANSS | 23 Apr 1992 | 6.1 | 115% | Green |
| Landers | ANSS | 28 Jun 1992 | 7.3 | 124% | Green |
| Eureka Valley | ANSS | 17 May 1993 | 6.1 | 92% | Yellow |
| Northridge | ANSS | 17 Jan 1994 | 6.7 | 107% | Yellow |
| Hector Mine | ANSS | 16 Oct 1999 | 7.1 | 103% | Yellow |
| San Simeon | ANSS | 22 Dec 2003 | 6.5 | 99% | Yellow |
| Parkfield | ANSS | 28 Sep 2004 | 6 | 136% | Green |
| El Mayor Cucapah | ANSS | 4 Apr 2010 | 7.2 | 124% | Green |
| Tottori | JMA | 6 Oct 2000 | 7.3 | 96% | Yellow |
| Ryukyu | JMA | 18 Dec 2001 | 7.3 | 191% | Green |
| Chūetsu | JMA | 23 Oct 2004 | 6.8 | 126% | Green |
| Southern Romoi | JMA | 14 Dec 2004 | 6.1 | 112% | Green |
| Fukuoka | JMA | 20 Mar 2005 | 7 | 103% | Yellow |
| Chūetsu Offshore | JMA | 16 Jul 2007 | 6.8 | 128% | Green |
| Iwate | JMA | 13 Jun 2008 | 7.2 | 104% | Yellow |
| Awaji Island | JMA | 12 Apr 2013 | 6.3 | 150% | Green |
| Nagano (for) | JMA | 22 Nov 2014 | 6.7 | 105% | Yellow |
| Kumamoto (for) | JMA | 14 Apr 2016 | 6.5 | 88% | Red |
| Kumamoto | JMA | 15 Apr 2016 | 7.3 | 138% | Green |
| Fukushima | JMA | 21 Nov 2016 | 7.4 | 178% | Green |
| L'Aquila | Gasperini et al.[27] | 6 Apr 2009 | 6.3 | 139% | Green |
| Amatrice | Gasperini et al.[27] | 24 Aug 2016 | 6.2 | 83% | Red |
| Norcia | Gasperini et al.[27] | 30 Oct 2016 | 6.5 | 116% | Green |
| Denali | AEIC | 3 Nov 2002 | 7.6 | 145% | Green |
| Tohoku (for) | JMA | 9 March 2011 | 7.3 | 73% | Red |
| Tohoku | JMA | 11 March 2011 | 9 | 140% | Green |

The last row shows the Tohoku increase according to Tormann et al.[33]. AEIC, Alaska Earthquake Information Center (http://earthquake.alaska.edu); ANSS, Advanced National Seismic System (https://earthquake.usgs.gov/monitoring/anss/); for, foreshock.

of a subsequent large event. However, there are several limitations to FTLS. First, the number of cases that we are able to investigate is still limited, because magnitude-6 or greater earthquakes are rare in areas with excellent network coverage. Especially lacking are more cases of true positives. An important implication of our work is, therefore, that seismic networks around the globe must substantially upgrade and further automate their processing procedures and increase in station density. Being able to detect and process magnitude-2 and larger events consistently and almost in real time during a vigorous 'aftershock' sequence is a major challenge that very few networks master today. The imprint of a mainshock on the size distribution of aftershocks decays away within a few kilometres of the rupture plane[24], so it is critically important to achieve relative hypocentre accuracies of around 1–2 km. This requirement currently limits, for example, the analysis of the $M_w = 7.3$ foreshock preceding the Tohoku $M_w = 9$ mainshock, or the compilation of an analysis-based global earthquake catalogue for systematic testing. However, promising improvements in seismic

networks are on the way because, for example, template matching of all waveforms recorded against a large set of template events[42,43] is becoming computationally feasible. The Kumamoto case is especially important because it highlights that the FTLS approach can be applied within a few hours of the mainshock.

A further limitation is that we lack a better physics-based understanding and predictive modelling capability of precursory sequences. However, our hypothesis presents a new angle in which aftershock sequences can be understood and modelled. In addition, stimulated by our findings presented here, new laboratory-based, numerical and field-data-based studies will advance our understanding in the near future. There is also a clear need to test our hypothesis in a fully prospective sense—the gold standard of earthquake forecasting[44,45]. Such tests have been initiated and will take many years to complete with meaningful statistical power. We would advocate, however, that in regions of the world with sufficient network coverage, seismologists and energy managers should consider adopting our FTLS as additional information for decision-making during seismic crises.

## Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-019-1606-4.

1. Jordan, T. H. et al. Operational earthquake forecasting state of knowledge and guidelines for utilization. Ann. Geophys. **54**, 315–391 (2011).
2. Lippiello, E., Giacco, F., Marzocchi, W., Godano, C. & de Arcangelis, L. Statistical features of foreshocks in instrumental and ETAS catalogs. Pure Appl. Geophys. **174**, 1679–1697 (2017).
3. Reasenberg, P. A. & Jones, L. M. California aftershock hazard forecast. Science **247**, 345–346 (1990).
4. Roeloffs, E. & Goltz, J. The California earthquake advisory plan: a history. Seismol. Res. Lett. **88**, 784–797 (2017).
5. Field, E. H. et al. The potential uses of operational earthquake forecasting. Seismol. Res. Lett. **87**, 313–322 (2016).
6. Zechar, J. D., Marzocchi, W. & Wiemer, S. Operational earthquake forecasting in Europe: progress, despite challenges. Bull. Earthquake Eng. **14**, 2459–2469 (2016).
7. Ogata, Y. Statistical models for earthquake occurrences and residual analysis for point processes. J. Am. Stat. Assoc. **83**, 9–27 (1988).
8. Ogata, Y. Space-time point-process models for earthquake occurrences. Ann. Inst. Stat. Math. **50**, 379–402 (1998).
9. Gerstenberger, M. C., Wiemer, S., Jones, L. M. & Reasenberg, P. A. Real-time forecasts of tomorrow's earthquakes in California. Nature **435**, 328–331 (2005).
10. Parsons, T. et al. Stress-based aftershock forecasts made within 24 h postmain shock: expected north San Francisco Bay area seismicity changes after the 2014 $M = 6.0$ West Napa earthquake. Geophys. Res. Lett. **41**, 8792–8799 (2014).
11. Stein, R. Earthquake conversations. Sci. Am. **288**, 72–79 (2003).
12. Woessner, J. et al. A retrospective comparative forecast test on the 1992 Landers sequence. J. Geophys. Res. **116**, B05305 (2011).
13. van Stiphout, T., Wiemer, S. & Marzocchi, W. Are short-term evacuations warranted? Case of the 2009 L'Aquila earthquake. Geophys. Res. Lett. **37**, L06306 (2010).
14. Gulia, L., Tormann, T., Wiemer, S., Herrmann, M. & Seif, S. Short-term probabilistic earthquake risk assessment considering time-dependent b values. Geophys. Res. Lett. **43**, 1100–1108 (2016).
15. Brodsky, E. E. & Lay T. Recognizing foreshocks from the 1 April 2014 Chile earthquake. Science **344**, 700–702 (2014).
16. Bouchon, M., Durand, V., Marsan, D., Karabulut, H. & Schmittbuhl, J. The long precursory phase of most large interplate earthquakes. Nat. Geosci. **6**, 299–302 (2013).
17. Ellsworth, W. L. & Bulut, F. Nucleation of the 1999 Izmit earthquake by a triggered cascade of foreshocks. Nat. Geosci. **11**, 531–535 (2018).
18. Gutenberg, B. & Richter, C. F. Frequency of earthquakes in California. Bull. Seismol. Soc. Am. **34**, 185–188 (1944).
19. Ishimoto, M. & Iida, I. Observations of earthquakes registered with the microseismograph constructed recently. Bull. Earthquake Res. Inst. Univ. Tokyo **17**, 443–478 (1936).
20. Goebel, T. H. W., Schorlemmer, D., Becker, T. W., Dresen, G. & Sammis, C. G. Acoustic emissions document stress changes over many seismic cycles in stick-slip experiments. Geophys. Res. Lett. **40**, 2049–2054 (2013).
21. Amitrano, D. Brittle-ductile transition and associated seismicity: Experimental and numerical studies and relationship with the b value. J. Geophys. Res. **108**, 1–15 (2003).

22. Scholz, C. H. The frequency-magnitude relation of microfracturing in rock and its relation to earthquakes. *Bull. Seismol. Soc. Am.* **58**, 399–415 (1968).
23. Schorlemmer, D., Wiemer, S. & Wyss, M. Variations in earthquake-size distribution across different stress regimes. *Nature* **437**, 539–542 (2005).
24. Gulia, L. et al. The effect of a mainshock on the size distribution of the aftershocks. *Geophys. Res. Lett.* **45**, 13277–13287 (2005).
25. Helmstetter, A. Comparison of short-term and time-independent earthquake forecast models for southern California. *Bull. Seismol. Soc. Am.* **96**, 90–106 (2006).
26. Vannucci, G., Gasperini, P., Lolli, B. & Gulia, L. Fast characterization of sources of recent Italian earthquakes from macroseismic intensities. *Tectonophysics* **750**, 70–92 (2019).
27. Gasperini, P., Lolli, B. & Vannucci, G. Empirical calibration of local magnitude data sets versus moment magnitude in Italy. *Bull. Seismol. Soc. Am.* **103**, 2227–2246 (2013).
28. Wiemer, S. & Wyss, M. Mapping the frequency-magnitude distribution in asperities: an improved technique to calculate recurrence times? *J. Geophys. Res.* **102**, 15115–15128 (1997).
29. Japan Meteorological Agency. *JMA catalogue* http://www.data.jma.go.jp/svd/eqev/data/bulletin/index_e.html.
30. Earthquake Research Committee. *Evaluation of the 2016 Kumamoto Earthquakes* (ERC, 2016); https://www.jishin.go.jp/main/chousa/16may_kumamoto2/index-e.htm.
31. Nanjo, K.Z. & Yoshida, A. Anomalous decrease in relatively large shocks and increase in the p and b values preceding the April 16, 2016, M 7.3 earthquake in Kumamoto, Japan. *EPS* **69**, 13 (2017).
32. Omi, T. et al. Implementation of a real-time system for automatic aftershock forecasting in Japan. *Seismol. Res. Lett.* **90**, 242–250 (2019).
33. Tormann, T., Enescu, B., Woessner, J. & Wiemer, S. Randomness of megathrust earthquakes implied by rapid stress recovery after the Japan earthquake. *Nat. Geosci.* **8**, 152–158 (2015).
34. Gomber, J. Unsettled earthquake nucleation. *Nat. Geosci.* **11**, 463–464 (2018).
35. Mignan, A. The debate on the prognostic value of earthquake foreshocks: a meta-analysis. *Sci. Rep.* **4**, 4099 (2014).
36. Tape, C. et al. Earthquake nucleation and fault slip complexity in the lower crust of central Alaska. *Nat. Geosci.* **11**, 536–541 (2018); author correction 11, 615 (2018).
37. Okada, Y. Internal deformation due to shear and tensile faults in a half-space. *Bull. Seismol. Soc. Am.* **82**, 1018–1040 (1992).
38. European Commission. *A Strategy for Europe on Nutrition, Overweight and Obesity Related Health Issues. White Paper COM (2007)* **279** (European Commission, 2007); https://ec.europa.eu/health/ph_determinants/life_style/nutrition/documents/nutrition_wp_en.pdf.
39. Department of Health. *Information for a Healthy New York: Asthma Action Plan* (New York State Department of Health, 2002); www.health.state.ny.us/diseases/asthma/pdf/4850.pdf.
40. Bommer, J. J. et al. Control of hazard due to seismicity induced by a hot fractured rock geothermal project. *Eng. Geol.* **83**, 287–306 (2006).
41. Mignan, A., Broccardo, M., Wiemer, S. & Giardini, D. Induced seismicity closed-form traffic light system for actuarial decision-making during deep fluid injections. *Sci. Rep.* **7**, 13607 (2017).
42. Li, Z. F., Meier, M. A., Hauksson, E., Zhan, Z. W. & Andrews, J. Machine learning seismic wave discrimination: application to earthquake early warning. *Geophys. Res. Lett.* **45**, 4773–4779 (2018).
43. Shelly, D. R. A 15 year catalog of more than 1 million low-frequency earthquakes: tracking tremor and slip along the deep San Andreas Fault. *J. Geophys. Res. Solid Earth* **122**, 3739–3753 (2017).
44. Jordan, T. H. Earthquake predictability, brick by brick. *Seismol. Res. Lett.* **77**, 3–6 (2006).
45. Schorlemmer, D. et al. The collaboratory for the study of earthquake predictability: achievements and priorities. *Seismol. Res. Lett.* **89**, 1305–1313 (2018).
46. Wells, D. L. & Coppersmith, K. J. New empirical relationships among magnitude, rupture length, rupture width, rupture area, and surface displacement. *Bull. Seismol. Soc. Am.* **84**, 974–1002 (1994).
47. Shi, Y. & Bolt, B. A. The standard error of the magnitude-frequency *b* value. *Bull. Seismol. Soc. Am.* **72**, 1677–1687 (1982).
48. Dziewonski, A. M., Chou, T. A. & Woodhouse, J. H. Determination of earthquake source parameters from waveform data for studies of global and regional seismicity. *J. Geophys. Res.* **86**, 2825–2852 (1981).
49. Ekström, G., Nettles, M. & Dziewoński, A. M. The global CMT project 2004–2010: centroid-moment tensors for 13,017 earthquakes. *Phys. Earth Planet. Inter.* **200–201**, 1–9 (2012).
50. Wiemer, S. & Wyss, M. Minimum magnitude of completeness in earthquake catalogs: examples from Alaska, the Western United States, and Japan. *Bull. Seismol. Soc. Am.* **90**, 859–869 (2000).
51. Woessner, J. & Wiemer, S. Assessing the quality of earthquake catalogues: estimating the magnitude of completeness and its uncertainty. *Bull. Seismol. Soc. Am.* **95**, 684–698 (2005).
52. Tormann, T., Wiemer, S. & Mignan, A. Systematic survey of high-resolution b value imaging along Californian faults: inference on asperities. *J. Geophys. Res.* **119**, 2029–2054 (2014).

## METHODS

The method that we propose for real-time discrimination between foreshock and aftershock sequences and that we apply to the Kumamoto and Amatrice–Norcia sequences is composed of the following steps.

**Selecting events near the rupture plane of the initiating event.** (1) The focal mechanism[48,49], which is available within minutes to hours after the origin time of a moderate-to-large initiating earthquake (here we consider a magnitude of 6 or larger), provides the information required to build a first-order source model: the magnitude, strike, dip and rake of the two possible nodal planes. Here we use the focal mechanism provided by the global centroid moment tensor solution because it is harmonized and available for all events studied (but other solutions are equally possible). Using the empirical formulas of Wells and Coppersmith[46], we model the two nodal planes: the length and width are derived directly from the magnitude and the rake, the three-dimensional orientation is given by strike and dip (see ref. [46] and our code presented here for the relevant equations for normal, strike-slip and reverse regimes or the general case).

(2) We then adjust the hypocentre of the initiating event to the one reported by the local network, because global centroid moment tensor hypocentres are much less accurate and we need to ensure consistency between after- and foreshocks and the initiating event.

(3) To determine the actual fault plane automatically, we select all events recorded in the sequences within three kilometres of each of the nodal planes and then choose the plane where most of the supposed aftershocks are located in. This we assume to be the source volume, referred to from here on as 'the box'. Typically, one to several hours of aftershocks are sufficient to select the right plane, and rapid source-inversion approaches can also deliver a finite fault model within 1–2 h.

**Constructing the time series.** (4) We divide the dataset into two parts: a pre- and a post-initiating-event catalogue. The start time of the pre-event catalogue depends on the quality and completeness of the local network and sometimes on avoiding overlap with past sequences (in our case, we choose 1 January 2012 for both Japan and Italy; in Italy, to avoid overlap with the L'Aquila aftershocks and in Kumamoto to avoid the influence of the 2011 Tohoku $M_w = 9$ megathrust event). The pre-event period should ideally contain several years of seismicity for a robust estimate. The post-event catalogue is then updated as new events emerge; in our case we analyse the subsequent two years of aftershocks.

(5) The two sub-catalogues are cut at magnitude 1, and then we compute the overall $M_c$ using the maximum-curvature method[50]. This defines the overall minimum $M_c$ level needed to make the sample-specific $M_c$ assessment more robust.

(6) Next, we estimate a pre-event reference $b$ value. We distinguish two cases, depending on the abundance of the events within the box:

(a) If more than a user-defined minimum number of events ($N_{pre}$) are available, we compute a time series. This is done by first re-assessing completeness for the first sample of 250 events using the maximum-curvature method but adding a correction factor of $+0.2$ (as recommended by Woessner and Wiemer[51]). As additional quality assurance steps, we require at least 50 events above completeness and also check if the sample passes the linearity test described in Tormann et al.[52]. The $b$ and $a$ values and their respective uncertainties are computed using a maximum-likelihood assessment[47]. The window is then moved forward by one event and the background reference $b$ value is computed as the median of all individual $b$ values in this time series.

(b) If fewer than $N_{pre}$ events are available, we use the $N_{pre}$ events that are nearest to the epicentre and then compute a single regional background $b$ value as

reference, following the computational approach defined in (a). This procedure was used for the $M_w = 6.5$ Kumamoto event (Fig. 2a), sampling a distance of up to 17 km from the epicentre.

(7) We estimate the post-event time series of $b$ values. We first remove the events recorded in the initial part of the sequence, which is typically highly incomplete and heterogeneous. This exclusion period depends on the quality of the seismic network and is an expert's choice. We then compute a time series of $b$ values again by the approach described in (a); however, we use a sample size of $N_{post} = 400$ to increase robustness and because aftershock sequences are very data-rich. We plot the time series and its uncertainty in Figs. 1, 2 and compute the per cent change with respect to the reference $b$ value. If the difference exceeds $+10\%$ or $-10\%$, we assign a traffic light colour of green or red, respectively; otherwise we assign orange (Fig. 3).

(8) The procedure described in (7) is repeated after the second mainshocks.

The main free parameters in our analysis are $N_{pre}$ and $N_{post}$. We tested that the results of our analysis do not critically depend on the choice of these parameters within reasonable ranges (for example, $N_{pre} = 150$–$300$, $N_{post} = 250$–$500$).

**Mapping $b$-value changes.** As additional information, we assess the spatial footprint of $b$-value changes (Figs. 1, 2g, h). The $b$-value maps are computed using ZMAP 7.0 (available at http://www.seismo.ethz.ch/en/research-and-teaching/products-software/software/ZMAP/) and post-processed using Matlab and Generic Mapping Tool. The relevant input parameters are:

(I) Background $b$-value map. On a regularly spaced grid of 2 km, the closest 250 events above the pre-cut $M_c$ of 1.0 are sampled, up to a maximum radius of 15 km. The node-specific $M_c$ is then estimated by the maximum-curvature method[51], by adding a 0.2 correction[50]. The $b$ values are computed using the maximum-likelihood method[47] (Figs. 1, 2f).

(II) Post-initiating-event maps. The same procedure is applied, but we use 400 events and add a correction of 0.4 to account for the more heterogeneous data. For these two intervals, we plot the per cent difference in $b$ value with respect to the background (Figs. 1, 2g, h).

## Data and code availability

The datasets generated and analysed during the current study, as well as the Matlab codes written for the analysis, are available at https://doi.org/10.3929/ethz-b-000357449.

**Author contributions** L.G. and S.W. conceived the analysis method and wrote the paper. L.G. performed the data analysis and created the figures.

# ARTICLE

# Dynamics and genomic landscape of CD8+ T cells undergoing hepatic priming

Alexandre P. Bénéchet[1,11], Giorgia De Simone[1,2,11], Pietro Di Lucia[1], Francesco Cilenti[2,3], Giulia Barbiera[3], Nina Le Bert[4], Valeria Fumagalli[1,2], Eleonora Lusito[3], Federica Moalli[1], Valentina Bianchessi[2,3], Francesco Andreata[1], Paola Zordan[1], Elisa Bono[1], Leonardo Giustini[1], Weldy V. Bonilla[5], Camille Bleriot[6], Kamini Kunasegaran[4], Gloria Gonzalez-Aseguinolaza[7], Daniel D. Pinschewer[5], Patrick T. F. Kennedy[8], Luigi Naldini[2,3], Mirela Kuka[1,2], Florent Ginhoux[6,9], Alessio Cantore[2,3], Antonio Bertoletti[4,6], Renato Ostuni[2,3,12], Luca G. Guidotti[1,2,12] & Matteo Iannacone[1,2,10,12]*

The responses of CD8+ T cells to hepatotropic viruses such as hepatitis B range from dysfunction to differentiation into effector cells, but the mechanisms that underlie these distinct outcomes remain poorly understood. Here we show that priming by Kupffer cells, which are not natural targets of hepatitis B, leads to differentiation of CD8+ T cells into effector cells that form dense, extravascular clusters of immotile cells scattered throughout the liver. By contrast, priming by hepatocytes, which are natural targets of hepatitis B, leads to local activation and proliferation of CD8+ T cells but not to differentiation into effector cells; these cells form loose, intravascular clusters of motile cells that coalesce around portal tracts. Transcriptomic and chromatin accessibility analyses reveal unique features of these dysfunctional CD8+ T cells, with limited overlap with those of exhausted or tolerant T cells; accordingly, CD8+ T cells primed by hepatocytes cannot be rescued by treatment with anti–PD-L1, but instead respond to IL-2. These findings suggest immunotherapeutic strategies against chronic hepatitis B infection.

Priming of circulating naive CD8+ T cells in non-lymphoid organs is hindered by the endothelial barrier that limits antigen recognition on epithelial cells. The liver is an exception: slow blood flow[1], the presence of endothelial fenestrations and the absence of a basement membrane allow CD8+ T cells to sense complexes of antigen and major histocompatibility complex (MHC) on hepatocytes[2,3]. Liver priming is thought to result in T cell unresponsiveness or dysfunction[4,5] but the underlying mechanisms, particularly in the context of hepatitis B virus (HBV) pathogenesis, are incompletely understood. HBV is a non-cytopathic virus that replicates in hepatocytes and causes acute or chronic infections[6,7]. Infection outcome is determined mainly by the kinetics, breadth, vigour and effector functions of HBV-specific CD8+ T cell responses[6]. Chronic HBV infection is typically acquired at birth or in early childhood[8] and proceeds from an initial 'immune-tolerant' phase (characterized by high viraemia and no liver inflammation) to an 'immune-active' phase (in which viraemia is lower and liver inflammation is present)[8,9]. HBV-specific CD8+ T cells in young, immune-tolerant patients are considered akin to exhausted T cells that characterize the immune-active phase[10], as well as to other infection- or cancer-related conditions of immune dysfunction, although a detailed characterization is lacking[11].

## Dynamics of naive CD8+ T cells after hepatic priming
To study the immune mechanisms of early HBV unresponsiveness, we initially analysed HBV-specific CD8+ T cells undergoing priming in a non-inflamed liver. In accordance with previous data[12], envelope-specific naive CD8+ T cell receptor (TCR) transgenic T cells (referred to as Env28 $T_N$ cells)[12] adoptively transferred into HBV replication-

competent transgenic mice expressing all viral proteins in the hepatocyte[13] proliferated but did not develop IFNγ-producing or cytolytic capacities (Extended Data Fig. 1a–d). As an effective CD8+ T cell response is induced in immunocompetent individuals exposed to HBV in adulthood[14], it remains to be determined whether this is due to cross-priming events in secondary lymphoid organs or whether the liver itself can support full effector differentiation.

Using a system in which T cell priming is restricted to the liver (Fig. 1a and Extended Data Fig. 1e–h), we injected naive CD8+ TCR transgenic T cells specific for the core protein of HBV (referred to as Cor93 $T_N$ cells)[12] into major urinary protein (MUP)-core transgenic mice[15], which exclusively express a non-secretable version of the HBV core protein in 100% of hepatocytes (Extended Data Fig. 1i). Two additional groups of mice served as controls (Fig. 1a): (1) wild-type mice, and (2) wild-type mice transduced with recombinant replication-defective, lymphocytic choriomeningitis virus (LCMV)-based vectors[16] that target a non-secretable version of the HBV core protein (known as rLCMV-core) to Kupffer cells and hepatic dendritic cells that are not naturally infected by HBV (Extended Data Fig. 1i). Antigen recognition was restricted to hepatocytes in MUP-core mice or to Kupffer cells and hepatic dendritic cells in rLCMV-transduced wild-type mice, as Cor93 $T_N$ cells isolated 1 h after transfer upregulated CD69 (a proxy for antigen recognition) in the liver but not in the blood, lung and bone marrow (Extended Data Fig. 1j). We then characterized the fate and function of naive CD8+ T cells undergoing intrahepatic priming. HBV-specific naive CD8+ T cells that recognize antigen in the liver underwent local activation (Extended Data Fig. 1j) and proliferation, so that by day 3 after transfer we could

**Fig. 1 | Spatiotemporal dynamics of naive CD8+ T cells undergoing intrahepatic priming. a**, Schematic of the experimental setup. Five million naive CD8+ T cells from Cor93 TCR transgenic mice (Cor93 $T_N$) were transferred into C57BL/6 (wild-type, WT) or MUP-core recipients. Mice were splenectomized and treated with anti-CD62L antibodies 48 h and 4 h before cell transfer, respectively. When indicated, mice were injected with $2.5 \times 10^5$ infectious units of non-replicating rLCMV-core 4 h before CD8+ T cell transfer. Livers were collected and analysed at the indicated time points. **b**, Absolute numbers of Cor93 T cells in the livers of indicated mice at indicated time points. $n$ (4 h and day 3) = 4 (WT), 7 (WT + rLCMV-core and MUP-core); $n$ (day 5) = 3 (WT), 13 (WT + rLCMV-core and MUP-core); $n$ (day 7) = 4 (WT), 6 (WT + rLCMV-core), 10 (MUP-core) mice. **c**, Frequency of IFNγ-producing Cor93 T cells in the livers of indicated mice at indicated time points. $n$ (4 h, days 3 and 7) = 3; $n$ (day 5) = 3 (WT), 6 (WT + rLCMV-core), 7 (MUP-core) mice. **d**, Mean fluorescent intensity (MFI) of PD-1 expression on Cor93 T cells in the livers of indicated mice. $n$ = 3 mice. **e**, Left, representative confocal immunofluorescence micrographs of liver sections from wild-type + rLCMV-core mice (top) or MUP-core mice (bottom) 3 days after transfer of Cor93 $T_N$ cells. Distribution of Cor93 T cells (green) relative to portal tracts (red) is shown. Sinusoids are in white. Scale bars, 100 μm. Middle, haematoxylin and eosin (H&E) staining of liver sections from the same mice. Dotted lines denote leukocyte clusters. Scale bars, 300 μm. Right, snapshots from representative intravital multiphoton microscopy movies of the same mice. Cor93 T cells tracks are in yellow, and blood vessels are in white. Scale bars, 40 μm. **f**, Mean speed of Cor93 T cells in the livers of indicated mice. $n$ = 613 tracks (WT + rLCMV-core), 156 tracks (MUP-core). **g**, Cor93 and Env28 naive CD8+ T cells were co-transferred into splenectomized and anti-CD62L-treated C57BL/6 × BALB/c $F_1$ (WT) or MUP-core × BALB/c $F_1$ (MUP-core) recipients. When indicated, mice were injected with rLCMV-env or rLCMV-core/env. Livers were collected and analysed 5 days after T cell transfer. Absolute numbers of total (left) and IFNγ-producing (right) Cor93 and Env28 T cells in the livers of indicated mice are indicated. $n$ = 4 mice. Data are mean ± s.e.m. and representative of at least three independent experiments. *$P < 0.05$, **$P < 0.01$, ***$P < 0.001$, two-tailed $t$-test (**b**, **c**, **f**) or one-way ANOVA with Bonferroni post-test (**d**, **g**).

recover approximately 30-fold more intrahepatic Cor93 T cells in antigen-expressing mice than in control mice (Fig. 1b). Whereas antigen recognition on Kupffer cells and hepatic dendritic cells yielded bona fide effector cells endowed with IFNγ-producing (Fig. 1c) and cytolytic abilities (data not shown), antigen recognition on hepatocytes led to the generation of dysfunctional cells that produced little or no IFNγ after in vitro peptide re-stimulation (Fig. 1c), did not develop cytotoxic activity, and instead upregulated the inhibitory receptor PD-1 (Fig. 1d). Together, these results indicate that—depending on the nature of the antigen-presenting cell—the liver can support the development of either functional or dysfunctional CD8+ T cells. Spatiotemporal analyses of mice transduced with rLCMV-core revealed T cell clusters scattered throughout the liver lobule (Fig. 1e and Extended Data Fig. 2a, b) in a pattern that is reminiscent of that observed during acute self-limited HBV infection[17]. By contrast, CD8+ T cells formed clusters that coalesced around portal tracts in MUP-core mice (Fig. 1e and Extended Data Fig. 2a, b)—a situation that is similar to chronic HBV infection[18]. These periportal clusters occurred despite the fact that the core protein is uniformly expressed in all hepatocytes[15] (Extended Data Fig. 1i) and that in the first few hours after transfer, CD8+ T cells recognize antigen on hepatocytes that can be distant from portal tracts (Extended Data Fig. 2). Multiphoton intravital imaging of the liver showed that the clusters formed in wild-type mice transduced with rLCMV-core are dense, extravascular and composed of largely immotile cells; by contrast, clusters formed in MUP-core mice are looser, intravascular and composed of more motile cells (Fig. 1e, f and Supplementary Videos 1, 2). By day 5–7, clusters in wild-type mice transduced with rLCMV-core start to disaggregate as cells move out from the liver, whereas clusters in MUP-core mice remain in place, possibly reflecting antigen persistence (Supplementary Video 3).

The notion that the liver can support full effector differentiation is not without precedent[4,19,20] but stands in contrast to the immunological dogma that T cell priming occurs exclusively in secondary lymphoid organs. As rLCMV targets both Kupffer cells and hepatic dendritic cells, we next investigated which of these two cell populations supports intrahepatic priming of naive CD8+ T cells. To this end, wild-type mice were injected with clodronate liposomes that effectively deplete Kupffer cells while sparing hepatic dendritic cells[21] (Extended Data Fig. 3a–c). As shown in Extended Data Fig. 3d–f, depletion of Kupffer cells abolished $T_N$ cell expansion and effector differentiation. We then depleted hepatic dendritic cells by injection of diphtheria toxin in wild-type mice reconstituted with CD11c-DTR bone marrow, but this treatment did not affect the ability of rLCMV-core to efficiently prime and promote effector differentiation of Cor93-specific CD8+ T cells (Extended Data Fig. 3g–k). Together, the data indicate that Kupffer cells—but not hepatic dendritic cells—promote effective CD8+ T cell priming on rLCMV injection.

We next evaluated the fate of naive T cells that are primed within livers expressing low levels of HBV core antigen (HBcAg). First, we transferred Cor93 $T_N$ cells into wild-type mice previously injected with a low dose of a hepatotropic adeno-associated viral vector (AAV) encoding the HBV core protein. This dose (transducing less than 5% of hepatocytes) supported Cor93 CD8+ T cell proliferation but not effector differentiation (Extended Data Fig. 4a–e). Second, we transferred Cor93 $T_N$ cells into 3–4-week-old MUP-core mice, which express only trace amounts of this protein per hepatocyte (core protein expression in these mice is developmentally regulated, reaching plateaus at 6–8 week of age[15]). As shown in Extended Data Fig. 4f–l, reducing the amount of expressed antigen by more than 15-fold within individual hepatocytes did not affect the differentiation of intrahepatically primed CD8+ T cells. Together, these experiments indicate that low expression of hepatocellular core antigen is per se not sufficient to induce effector differentiation.
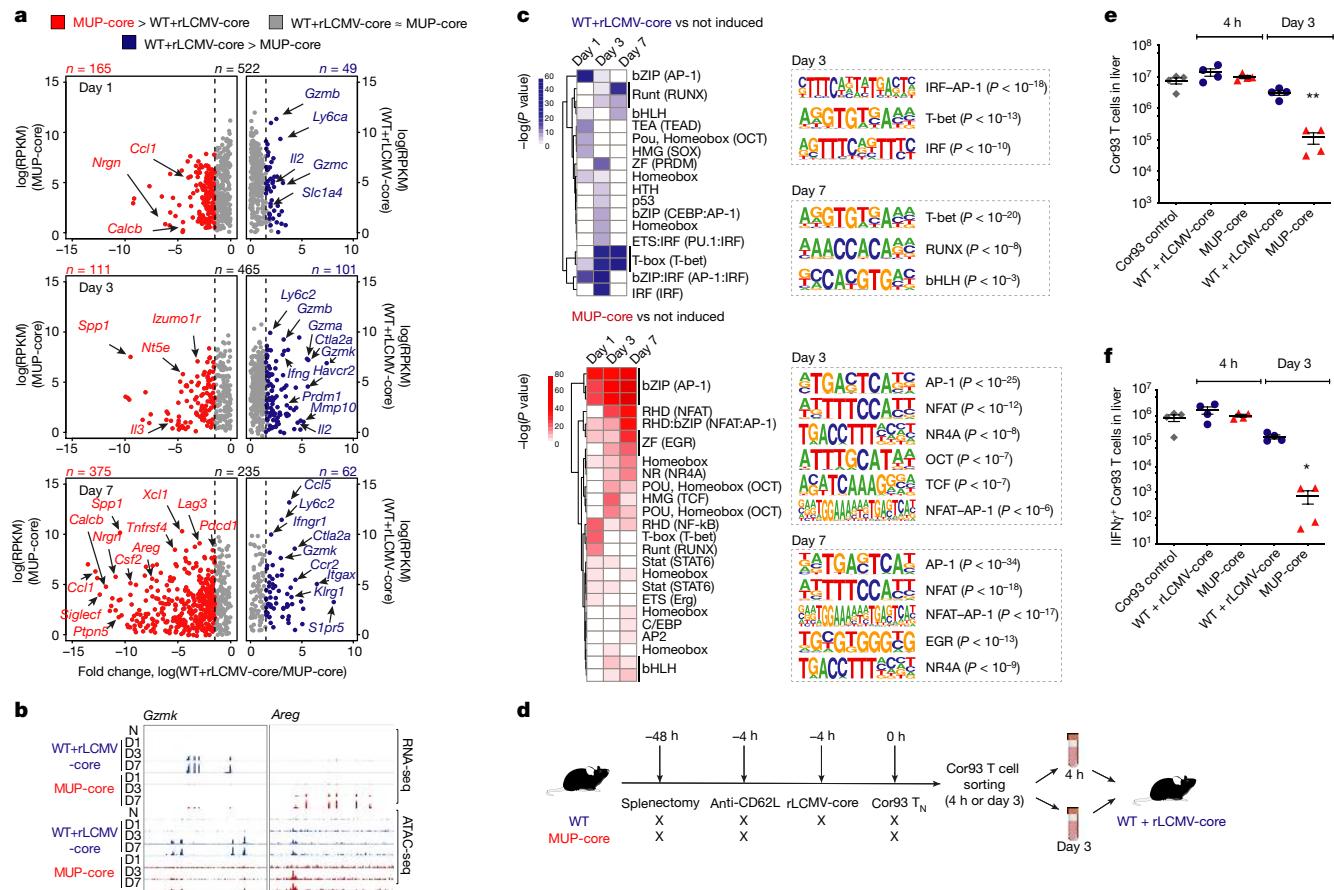
**Fig. 2 | Transcriptomic and chromatin accessibility analyses of CD8+ T cells undergoing intrahepatic priming. a**, Scatter plot showing the level ($y$ axis) and difference in expression ($x$ axis) of inducible genes in the dataset (versus Cor93 $T_N$) in the indicated conditions. Genes expressed at higher levels in WT + rLCMV-core or MUP-core mice are shown in blue or red, respectively. Naive ($n = 2$), WT + rLCMV-core ($n = 3$) and MUP-core (day 1 and 3, $n = 2$; day 7, $n = 3$). Differential gene expression was evaluated fitting a negative binomial generalized linear model on the dataset and then performing a quasi-likelihood $F$-test. The Benjamini–Hochberg procedure was applied to correct for multiple tests. **b**, Integrative Genomics Viewer (IGV)[40] snapshots showing RNA-seq and ATAC-seq data at *Gzmk* and *Areg* loci, selected as representative genes with differential expression in WT + rLCMV-core or MUP-core mice, respectively. D, day; N, naive. **c**, Left, heat map showing the enrichment of DNA motifs (HOMER)[41] within the top 200 inducible (versus Cor93 $T_N$) and differential ATAC-seq peaks in WT + rLCMV-core (blue) or MUP-core (red) mice. A set of 3,899 non-inducible ATAC-seq peaks was used as background. Right, selected enriched motifs and putative cognate

transcription factors in ATAC-seq peaks from WT + rLCMV-core (top) or MUP-core (bottom) mice. Motif enrichment was calculated using cumulative binomial distributions. Naive ($n = 2$), WT + rLCMV-core (day 1 and 7, $n = 2$; day 3, $n = 3$) and MUP-core (day 1 and 3, $n = 2$; day7, $n = 3$). **d**, Schematic of the experimental setup. Five million Cor93 $T_N$ cells were transferred into wild-type or MUP-core recipients. Mice were splenectomized and treated with anti-CD62L 48 h and 4 h before cell transfer, respectively. Where indicated, mice were injected with $2.5 \times 10^5$ infectious units of non-replicating rLCMV-core 4 h before Cor93 $T_N$ cell transfer. Livers were collected either 4 h or 3 days after cell transfer. Then, $5 \times 10^3$ purified Cor93 T cells were injected back into rLCMV-core-injected wild-type mice (which were splenectomized and treated with anti-CD62L as previously described). Livers were collected and analysed by flow cytometry 5 days after Cor93 T cell transfer. **e, f**, Absolute numbers of total (**e**) and IFNγ-producing (**f**) Cor93 T cells in the livers of the indicated mice. $n = 4$ mice. \*$P < 0.05$, \*\*$P < 0.01$, one-way ANOVA with Bonferroni post-test. Data are mean ± s.e.m. and representative of at least three independent experiments.

Finally, we investigated the fate of intrahepatic $T_N$ cells primed by antigen presented by both hepatocytes as well as Kupffer cells and dendritic cells by transferring Env28 and Cor93 $T_N$ cells into wild-type and MUP-core mice transduced with rLCMV vectors that encode either the HBV envelope protein alone (rLCMV-env) or both the HBV core and envelope proteins (rLCMV-core/env). As expected, in wild-type mice, $T_N$ cells expanded and differentiated into IFNγ-secreting cells only when cognate antigen was present (Fig. 1g). In MUP-core mice, the injection of rLCMV-env allowed for Env28 (but not Cor93) $T_N$ cell expansion and effector differentiation, which indicates that (i) innate immune signals carried by rLCMV vectors are not sufficient to overcome Cor93 T cell dysfunction; and (ii) dysfunctional Cor93 T cells do not produce soluble or membrane-bound mediators that inhibit Env28 T cell effector differentiation (Fig. 1g). Finally, injection of rLCMV-core/env to MUP-core mice led to Env28 (but not Cor93) $T_N$ cell expansion and effector differentiation, indicating that—when antigen is presented by both hepatocytes as well as Kupffer and hepatic

dendritic cells—hepatocellular antigen presentation is dominant in inducing immune dysfunction (Fig. 1g).

## Genomic analysis of CD8+ T cells after hepatic priming
To unveil molecular determinants of this immune dysfunction, we performed transcriptomic (RNA sequencing, RNA-seq) and chromatin accessibility (assay for transposase-accessible chromatin using sequencing, ATAC-seq) analyses of Cor93 CD8+ T cells isolated from the livers of control wild-type mice transduced with rLCMV-core or of MUP-core mice at days 1, 3 and 7 after transfer. We observed a broad and progressive transcriptional divergence in intrahepatic Cor93 CD8+ T cells sorted from the two groups of mice (Fig. 2a, Extended Data Fig. 5a and Supplementary Tables 1, 2). Hepatic CD8+ T cells from wild-type rLCMV-core-transduced mice, but not those from MUP-core mice, upregulated canonical genes of the T cell effector program such as *Gzma*, *Gzmb* and *Ifng*[22,23]. By contrast, CD8+ T cells isolated from the livers of MUP-core mice upregulated transcripts that encode a different
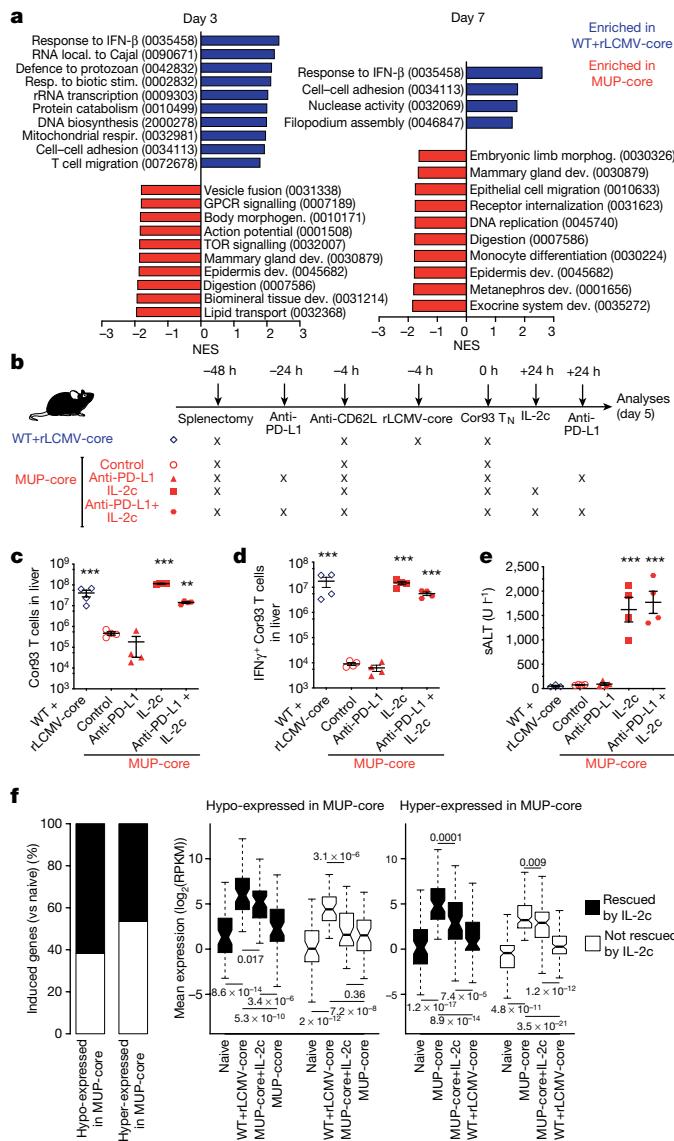
**Fig. 3 | Intrahepatically primed, dysfunctional CD8+ T cells can be rescued by IL-2, but not by anti-PD-L1 antibodies. a**, Normalized enrichment score (NES) of selected GO categories enriched within genes expressed at higher levels in WT + rLCMV-core (blue, positive values) or MUP-core (red, negative values) mice at the indicated time points. GO categories were identified by GSEA[42] and grouped by similarity with REVIGO[43]. **b**, Schematic of the experimental setup. **c, d**, Absolute numbers of total (**c**) and IFNγ-producing (**d**) Cor93 T cells in the livers of the indicated mice $n = 4$. **e**, Serum ALT (sALT) levels of the indicated mice. $n = 3$ (WT + rLCMV-core) or 4 (all other groups). **f**, Left, stacked bar plot showing the effect of IL-2c on genes induced at day 5 (versus naive) in Cor93 T cells from WT + rLCMV-core or MUP-core mice. Genes hypo-expressed or hyper-expressed in MUP-core mice as compared with WT + rLCMV-core mice are shown separately. Right, box plots showing expression levels of hypo-expressed (left) or hyper-expressed (right) genes at day 5 in the indicated conditions. Genes the expression of which is rescued or not rescued in MUP-core + IL-2c mice are shown in black or white, respectively. Naive ($n = 2$), WT + rLCMV-core ($n = 3$) and MUP-core (day 1 and 3, $n = 2$; day 7, $n = 3$). For all box plots, horizontal line denotes the median; lower and upper limits of the box represent the first and third quartile, respectively, and whiskers extend up to 1.5 times the interquartile range. Data in **c–e** are mean ± s.e.m. All data are representative of at least two independent experiments. $P$ values in **f** determined by two-sided Wilcoxon rank-sum test. **$P < 0.01$, ***$P < 0.001$, one-way ANOVA with Bonferroni post-test (**c–e**).

set of cytokines and chemokines (*Ccl1*, *Csf2* and *Xcl1*), growth factors and hormones (*Areg* and *Calcb*), inhibitory molecules (*Pdcd1*, *Lag3* and *Havcr2*) or surface markers (*Siglecf*) (Fig. 2a, b and Supplementary

Table 2). CD8+ T cells from wild-type rLCMV-core-transduced mice or from MUP-core mice had distinct chromatin accessibility profiles at days 3 and 7 after transfer (Extended Data Fig. 5b, c and Supplementary Table 3). Motif enrichment analysis on differentially induced (versus naive CD8+ T cells) ATAC-seq peaks revealed an over-representation of binding sites for transcription factor families involved in effector T cell differentiation, such as IRF, IRF–AP-1 and T-bet at day 3, as well as T-bet, RUNX and bHLH at day 7 in CD8+ T cells from wild-type rLCMV-core-transduced mice[24–27]. By contrast, ATAC-seq peaks of CD8+ T cells from MUP-core mice were enriched in binding sites for AP-1, NFAT, NFAT–AP-1 as well as for NR4A (recently associated with CD8+ T cell dysfunction[28,29]), OCT, TCF and EGR (Fig. 2c, Supplementary Table 4).

Our genomic analyses indicated that antigen recognition on Kupffer cells can support priming and differentiation into effector CD8+ T cells similar to those recovered from secondary lymphoid organs (Supplementary Table 5). By contrast, antigen recognition on hepatocytes initiates a defective differentiation program with progressive accumulation of chromatin and transcriptional landscape alterations that ultimately underlie a dysregulated T cell phenotype.

We next looked at the plasticity of the dysfunctional Cor93 T cells recovered from MUP-core livers. When Cor93 T cells were sorted from MUP-core livers 4 h after injection and then transferred into wild-type rLCMV-core-transduced mice, they were fully capable of expanding and differentiating into effector cells (Fig. 2d–f). By contrast, Cor93 T cells isolated from MUP-core livers at day 3 (a time point in which chromatin alterations are evident) (Fig. 2c and Extended Data Fig. 5) and transferred into wild-type rLCMV-core-transduced mice were significantly impaired in their ability to expand and differentiate into IFNγ-producing cells (Fig. 2d–f). These data indicate that three days of hepatocellular antigen exposure are sufficient to render cells partially refractory to effector differentiation stimuli.

Gene set enrichment analysis (GSEA) identified distinct sets of Gene Ontology (GO) categories in the transcriptomes of CD8+ T cells from the two groups. Genes with higher expression in CD8+ T cells from wild-type mice transduced with rLCMV-core were enriched in GO categories linked to effector immune responses such as responses to type I interferon, cell proliferation, T cell migration and cell–cell adhesion. By contrast, CD8+ T cells from livers of MUP-core mice did not express genes linked to effector T cell responses beyond day 1, and instead expressed genes belonging to GO categories linked to tissue development and organ remodelling, cell differentiation and cell-matrix interaction (Fig. 3a, Extended Data Fig. 6 and Supplementary Table 6). The transcriptional program of hepatic CD8+ T cells isolated from MUP-core mice at day 7 after transfer was not obviously overlapping with that of other known dysfunctional CD8+ T cell fates, as genes with selective expression in these cells were poorly expressed in reference transcriptomic datasets of splenic LCMV-specific exhausted CD8+ T cells[30,31] or tolerant self-antigen-specific CD8+ T cells[32] (Extended Data Fig. 7 and Supplementary Tables 7, 8). An exhaustion-like signature[30], however, was progressively enriched in the transcriptome of CD8+ T cells from MUP-core mice at days 3 and 7 after transfer, as determined by GSEA (Extended Data Fig. 7). These data indicate that, while priming by hepatocytes initiates a unique dysfunctional program, hepatocellular antigen persistence gradually triggers an additional exhaustion profile.

## Dysfunctional CD8+ T cells rescued by IL-2 treatment

Among the genes that are differentially expressed (Fig. 2a), we focused on two known regulators of T cell function: *Pdcd1* and *Il2*[33–36]. *Pdcd1* was hyper-expressed in hepatic Cor93 CD8+ T cells sorted from MUP-core mice (Fig. 2a), whereas *Il2* was found to be induced in the livers of wild-type mice transduced with rLCMV-core as well as hyper-expressed on Cor93 CD8+ T cells sorted from the livers of wild-type mice transduced with rLCMV (Fig. 2a). We assessed the functional consequences of these findings by treating Cor93 TN-cell-injected MUP-core mice with anti-PD-L1 blocking antibodies, with recombinant IL-2 coupled with anti-IL-2 antibodies (IL-2c)[37], or with
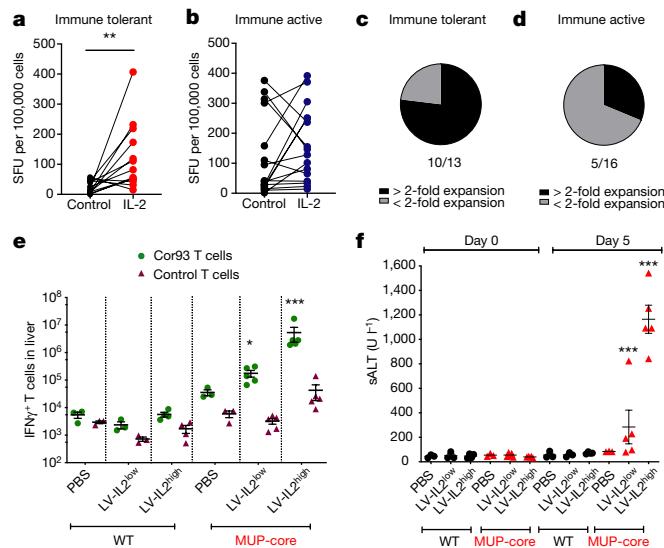
**Fig. 4 | Therapeutic potential of IL-2 treatment for restoration of T cells during chronic HBV infection. a, b,** HBV-specific T cell frequency from 13 immune-tolerant (**a**) and 16 immune-active (**b**) patients with chronic HBV (Supplementary Table 10) cultured with or without IL-2. **c, d,** The percentage of immune-tolerant (**c**) and immune-active (**d**) patients for which the HBV-specific T cell expansion increased by more than twofold after the addition of IL-2 are shown in black. **e,** Absolute numbers of IFN$\gamma$-producing T cells in the livers of the indicated mice. **f,** Serum ALT levels at days 0 and 5 in the same mice. In **e** and **f**, $n = 3$ (WT + PBS, WT + LV-IL-2$^{low}$ and MUP-core + PBS), 4 (WT + LV-IL-2$^{high}$) or 5 (MUP-core + LV-IL-2$^{low}$ and MUP-core + LV-IL-2$^{high}$) mice. *$P < 0.05$, ***$P < 0.001$, Wilcoxon matched-pairs signed-rank test (**a, b**) or one-way ANOVA with Bonferroni post-test (**e, f**). Data in **e** and **f** are mean ± s.e.m., and all data are representative of at least three independent experiments.

a combination of both (Fig. 3b). Administration of IL-2c promoted expansion and differentiation of Cor93 T cells into IFN$\gamma$-producing, cytotoxic effector cells (Fig. 3c–e), whereas anti-PD-L1 treatment either did not do so when given alone or did not show a synergistic effect when given in combination with IL-2c (Fig. 3c–e). Administration of IL-2c 1 day after transfer of Cor93 $T_N$ cells into MUP-core mice substantially rescued the transcriptional program of dysfunctional CD8$^+$ T cells, as measured by RNA-seq at day 5 (Fig. 3f, Extended Data Fig. 8 and Supplementary Table 9). More than half of the genes with defective expression (hypo-expressed genes) in hepatic CD8$^+$ T cells from MUP-core mice were upregulated in IL-2c-treated MUP-core mice, often reaching expression levels comparable to those detected in wild-type mice injected with rLCMV-core. Similarly, a comparable fraction of genes with higher expression (hyper-expressed genes) in hepatic CD8$^+$ T cells from MUP-core mice were downregulated by IL-2c treatment (Fig. 3f, Extended Data Fig. 8 and Supplementary Table 9).

To assess the specificity of our treatment[37,38], we co-transferred antigen-specific (Cor93) and irrelevant (Env28) $T_N$ cells into MUP-core mice 24 h before IL-2c administration. Cor93 and Env28 $T_N$ cells were also transferred into control wild-type mice previously injected with rLCMV-core/env. IL-2c improved the ability of antigen-specific Cor93 T cells to expand, differentiate into IFN$\gamma$-producing cells and accumulate in clusters scattered throughout the liver lobules, but it had no effect on irrelevant Env28 $T_N$ cells (data not shown).

## Therapeutic potential of IL-2

Next, we tested the effect of IL-2 treatment in HBV replication-competent transgenic mice that were neither splenectomized nor treated with anti-CD62L blocking antibodies. IL-2c administration promoted the differentiation of Cor93 T cells into IFN$\gamma$-producing, cytotoxic effector cells that accumulated in clusters scattered throughout the liver lobules and exerted potent antiviral activity (Extended Data Fig. 9).

We then hypothesized that HBV-specific T cells present in immune-tolerant patients have a different functional behaviour than those present in immune-active patients and might be more closely related to T cells primed by hepatocytes in the mouse models. Peripheral T cells from 13 immune-tolerant and 16 immune-active patients (Supplementary Table 10) were stimulated with overlapping HBV peptides in the presence or absence of recombinant human IL-2, and the frequency of HBV-specific T cells was determined by IFN$\gamma$ ELISpot assay. Only very low frequencies of IFN$\gamma$-secreting cells were detected in immune-tolerant patients in the absence of IL-2 (mean = 24 spot-forming units (SFU) per $10^5$ cells, Fig. 4a); the addition of IL-2, however, significantly augmented their frequency in 10 out of the 13 patients (mean = 122 SFU per $10^5$ cells; Fig. 4a, c). By contrast, HBV-specific T cells from immune-active patients did not require and could not be boosted by IL-2 during their expansion, and their frequency was similar to that of immune-tolerant patients in the presence of IL-2 (Fig. 4b, d). The data suggest that HBV-specific T cells from immune-tolerant, but not immune-active, patients resemble hepatocellularly primed mouse CD8$^+$ T cells in that they can expand and secrete IFN$\gamma$ only after IL-2 treatment. Whether IL-2 exerts an even greater effect on HBV-specific T cell restoration if administered directly to immune-tolerant patients (in which Kupffer cells could cross-present hepatocellular antigens) remains to be determined.

To test the clinical potential of IL-2 in a system that may limit its systemic toxicity[34], we generated third-generation, self-inactivating lentiviral vectors (LV.ET.mIL2.142T) that allow selective hepatocellular expression of mouse IL-2[39]. We injected wild-type or MUP-core mice with $2.5 \times 10^8$ (LV-IL2$^{low}$) or $5 \times 10^8$ (LV-IL2$^{high}$) transducing units per mouse, 7 days before Cor93 or control $T_N$ cell injection. As shown in Fig. 4e, f, lentiviral-mediated hepatic expression of IL-2, even at a dose that transduces less than 10% of hepatocytes in vivo, increased the capacity of Cor93 (but not control) T cells to expand and differentiate into IFN$\gamma$-producing cells endowed with cytolytic capacities.

## Discussion

We have delineated the spatiotemporal dynamics, genomic landscape and functional consequences of naive CD8$^+$ T cells undergoing intrahepatic priming (Extended Data Fig. 10). We showed that hepatocellular presentation leads to a CD8$^+$ T cell dysfunction that is distinct from T cell alterations reported in other viral infections and cancer and, as such, is not readily responsive to anti-PD-L1 treatment. As immune checkpoint inhibitors are beginning to be tested in patients persistently infected with HBV, the results reported here should help to interpret the outcome of those studies and eventually inform the design of modified trials in selected cohorts of patients. Our data identify IL-2 as a potent immunotherapeutic that can rescue CD8$^+$ T cells rendered dysfunctional by hepatocellular priming. Thus, IL-2-based strategies should be considered for the treatment of chronic HBV infection.

## Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-019-1620-6.

1. Sironi, L. et al. *In vivo* flow mapping in complex vessel networks by single image correlation. *Sci. Rep.* **4**, 7341 (2014).
2. Warren, A. et al. T lymphocytes interact with hepatocytes through fenestrations in murine liver sinusoidal endothelial cells. *Hepatology* **44**, 1182–1190 (2006).
3. Guidotti, L. G. et al. Immunosurveillance of the liver by intravascular effector CD8$^+$ T cells. *Cell* **161**, 486–500 (2015).
4. Wong, Y. C., Tay, S. S., McCaughan, G. W., Bowen, D. G. & Bertolino, P. Immune outcomes in the liver: is CD8 T cell fate determined by the environment? *J. Hepatol.* **63**, 1005–1014 (2015).

5. Holz, L. E. et al. Naive CD8 T cell activation by liver bone marrow-derived cells leads to a "neglected" IL-2[low] Bim[high] phenotype, poor CTL function and cell death. *J. Hepatol.* **57**, 830–836 (2012).
6. Guidotti, L. G. & Chisari, F. V. Immunobiology and pathogenesis of viral hepatitis. *Annu. Rev. Pathol.* **1**, 23–61 (2006).
7. Wieland, S. F. & Chisari, F. V. Stealth and cunning: hepatitis B and hepatitis C viruses. *J. Virol.* **79**, 9369–9380 (2005).
8. Kennedy, P. T. F., Litwin, S., Dolman, G. E., Bertoletti, A. & Mason, W. S. Immune tolerant chronic hepatitis B: the unrecognized risks. *Viruses* **9**, 96 (2017).
9. European Association for the Study of the Liver. EASL 2017 Clinical Practice Guidelines on the management of hepatitis B virus infection. *J. Hepatol.* **67**, 370–398 (2017).
10. Fisicaro, P. et al. Targeting mitochondrial dysfunction can restore antiviral activity of exhausted HBV-specific CD8 T cells in chronic hepatitis B. *Nat. Med.* **23**, 327–336 (2017).
11. Kennedy, P. T. F. et al. Preserved T-cell function in children and young adults with immune-tolerant chronic hepatitis B. *Gastroenterology* **143**, 637–645 (2012).
12. Isogawa, M., Chung, J., Murata, Y., Kakimi, K. & Chisari, F. V. CD40 activation rescues antiviral CD8+ T cells from PD-1-mediated exhaustion. *PLoS Pathog.* **9**, e1003490 (2013).
13. Guidotti, L. G., Matzke, B., Schaller, H. & Chisari, F. V. High-level hepatitis B virus replication in transgenic mice. *J. Virol.* **69**, 6158–6169 (1995).
14. Bertoletti, A. & Ferrari, C. Adaptive immunity in HBV infection. *J. Hepatol.* **64** (Suppl.), S71–S83 (2016).
15. Guidotti, L. G., Martinez, V., Loh, Y. T., Rogler, C. E. & Chisari, F. V. Hepatitis B virus nucleocapsid particles do not cross the hepatocyte nuclear membrane in transgenic mice. *J. Virol.* **68**, 5469–5475 (1994).
16. Flatz, L. et al. Development of replication-defective lymphocytic choriomeningitis virus vectors for the induction of potent CD8+ T cell immunity. *Nat. Med.* **16**, 339–345 (2010).
17. Guidotti, L. G. et al. Viral clearance without destruction of infected cells during acute HBV infection. *Science* **284**, 825–829 (1999).
18. Ishak, K. et al. Histological grading and staging of chronic hepatitis. *J. Hepatol.* **22**, 696–699 (1995).
19. Klein, I. & Crispe, I. N. Complete differentiation of CD8+ T cells activated locally within the transplanted liver. *J. Exp. Med.* **203**, 437–447 (2006).
20. Böttcher, J. P. et al. IL-6 trans-signaling-dependent rapid development of cytotoxic CD8+ T cell function. *Cell Reports* **8**, 1318–1327 (2014).
21. Sitia, G. et al. Kupffer cells hasten resolution of liver immunopathology in mouse models of viral hepatitis. *PLoS Pathog.* **7**, e1002061 (2011).
22. Wherry, E. J. et al. Molecular signature of CD8+ T cell exhaustion during chronic viral infection. *Immunity* **27**, 670–684 (2007).
23. Best, J. A. et al. Transcriptional insights into the CD8+ T cell response to infection and memory T cell formation. *Nat. Immunol.* **14**, 404–412 (2013).
24. Dominguez, C. X. et al. The transcription factors ZEB2 and T-bet cooperate to program cytotoxic T cell terminal differentiation in response to LCMV viral infection. *J. Exp. Med.* **212**, 2041–2056 (2015).
25. Intlekofer, A. M. et al. Effector and memory CD8+ T cell fate coupled by T-bet and eomesodermin. *Nat. Immunol.* **6**, 1236–1244 (2005).
26. Cruz-Guilloty, F. et al. Runx3 and T-box proteins cooperate to establish the transcriptional program of effector CTLs. *J. Exp. Med.* **206**, 51–59 (2009).
27. Kurachi, M. et al. The transcription factor BATF operates as an essential differentiation checkpoint in early effector CD8+ T cells. *Nat. Immunol.* **15**, 373–383 (2014).
28. Chen, J. et al. NR4A transcription factors limit CAR T cell function in solid tumours. *Nature* **567**, 530–534 (2019).
29. Liu, X. et al. Genome-wide analysis identifies NR4A1 as a key mediator of T cell dysfunction. *Nature* **567**, 525–529 (2019).
30. Scott-Browne, J. P. et al. Dynamic changes in chromatin accessibility occur in CD8+ T cells responding to viral infection. *Immunity* **45**, 1327–1340 (2016).
31. Pauken, K. E. et al. Epigenetic stability of exhausted T cells limits durability of reinvigoration by PD-1 blockade. *Science* **354**, 1160–1165 (2016).
32. Schietinger, A., Delrow, J. J., Basom, R. S., Blattman, J. N. & Greenberg, P. D. Rescued tolerant CD8 T cells are preprogrammed to reestablish the tolerant state. *Science* **335**, 723–727 (2012).
33. Sharpe, A. H. & Pauken, K. E. The diverse functions of the PD1 inhibitory pathway. *Nat. Rev. Immunol.* **18**, 153–167 (2018).
34. Spolski, R., Li, P. & Leonard, W. J. Biology and regulation of IL-2: from molecular mechanisms to human therapy. *Nat. Rev. Immunol.* **18**, 648–659 (2018).
35. Manske, K. et al. Outcome of anti-viral immunity in the liver is shaped by the level of antigen expressed in infected hepatocytes. *Hepatology* **68**, 2089–2105 (2018).
36. Tolksdorf, F. et al. The PDL1-inducible GTPase Arl4d controls T effector function by limiting IL-2 production. *Sci. Rep.* **8**, 16123 (2018).
37. Boyman, O., Kovar, M., Rubinstein, M. P., Surh, C. D. & Sprent, J. Selective stimulation of T cell subsets with antibody-cytokine immune complexes. *Science* **311**, 1924–1927 (2006).
38. Kamimura, D. & Bevan, M. J. Naive CD8+ T cells differentiate into protective memory-like cells after IL-2 anti IL-2 complex treatment *in vivo*. *J. Exp. Med.* **204**, 1803–1812 (2007).
39. Brown, B. D., Venneri, M. A., Zingale, A., Sergi Sergi, L. & Naldini, L. Endogenous microRNA regulation suppresses transgene expression in hematopoietic lineages and enables stable gene transfer. *Nat. Med.* **12**, 585–591 (2006).
40. Robinson, J. T. et al. Integrative genomics viewer. *Nat. Biotechnol.* **29**, 24–26 (2011).
41. Heinz, S. et al. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol. Cell* **38**, 576–589 (2010).
42. Subramanian, A. et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl Acad. Sci. USA* **102**, 15545–15550 (2005).
43. Supek, F., Bošnjak, M., Škunca, N. & Šmuc, T. REVIGO summarizes and visualizes long lists of gene ontology terms. *PLoS ONE* **6**, e21800 (2011).

## METHODS

**Data reporting.** No statistical methods were used to predetermine sample size. The experiments were not randomized, and investigators were not blinded to allocation during experiments and outcome assessment.

**Mice.** C57BL/6, CD45.1 (inbred C57BL/6), BALB/c, Thy1.1 (CBy.PL(B6)-*Thy$^a$*/ScrJ), β-actin-GFP (C57BL/6-Tg(CAG-EGFP)1Osb/J), β-actin-DsRed (B6. Cg-Tg(CAG-DsRed*MST)1Nagy/J), *Tap1*-deficient (B6.129S2-*Tap1$^{tm1Arp}$*/J), TCR-I (B6.Cg-Tg(TcraY1, TcrbY1)416Tev/J), CD11c-DTR (B6.FVB-*1700016L2Rik$^{Tg(Itgax-DTR/EGFP)57Lan}$*/J) mice were purchased from Charles River or The Jackson Laboratory. MHC-II$^{-/-}$ mice were obtained through the Swiss Immunological Mutant Mouse Repository. MUP-core transgenic mice (lineage MUP-core 50 (MC50), inbred C57BL/6, H-2$^b$), that express the HBV core protein in 100% of the hepatocytes under the transcriptional control of the mouse *Mup1* promoter, have previously been described[15]. HBV replication-competent transgenic mice (lineage 1.3.32, inbred C57BL/6, H-2$^b$), that express all of the HBV antigens and replicate HBV in the liver at high levels without any evidence of cytopathology, have previously been described[13]. In indicated experiments, MUP-core and HBV replication-competent transgenic mice were used as C57BL/6 × BALB/c H-2$^{bxd}$ F$_1$ hybrids. Cor93 TCR transgenic mice (lineage BC10.3, inbred CD45.1), in which >98% of the splenic CD8$^+$ T cells recognize a K$^b$-restricted epitope located between residues 93 and 100 in the HBV core protein (MGLKFRQL), have previously been described[12]. Env28 TCR transgenic mice (lineage 6C2.36, inbred Thy1.1 BALB/c), in which approximately 83% of the splenic CD8$^+$ T cells recognize a L$^d$-restricted epitope located between residues 28 and 39 of the HBV surface antigen (HBsAg; IPQSLDSWWTSL), have previously been described[12]. For imaging experiments, Cor93 and TCR-I transgenic mice were bred against both β-actin-GFP and β-actin-DsRed mice, and Env28 transgenic mice were bred against β-actin-DsRed mice that were previously backcrossed more than 10 generations against BALB/c. Bone marrow chimaeras were generated by irradiation of MUP-core or C57BL/6 mice with one dose of 9 Gy and reconstitution with the indicated bone marrow; mice were allowed to reconstitute for at least 8 weeks before use. In some experiments, to achieve full reconstitution of Kupffer cells from donor-derived bone marrow, mice were injected with 200 μl of clodronate-containing liposomes 28 and 31 days after bone marrow injection. Mice were housed under specific-pathogen-free conditions and used at 8–10 weeks of age, unless otherwise indicated. In all experiments, mice were matched for age, sex and (for the 1.3.32 mice) levels of serum HBV e-antigen (HBeAg) before experimental manipulations. In selected experiments, 1.3.32 mice were matched for serum levels of HBV DNA before experimental manipulations. All experimental animal procedures were approved by the Institutional Animal Committee of the San Raffaele Scientific Institute and are compliant with all relevant ethical regulations.

**Viruses and viral vectors.** Replication-incompetent LCMV-based vectors encoding HBV core protein, HBV envelope protein, HBV core and envelope proteins, or Cre recombinase (termed rLCMV-core, rLCMV-env, rLCMV-core/env and rLCMV-cre, respectively) were generated, grown and titrated as previously described[16]. Mice were injected intravenously with $2.5 \times 10^5$ infectious units of the indicated rLCMV vector 4 h before CD8$^+$ T cell injection.

Adeno-associated viruses expressing GFP and HBV core protein (AAV-core-GFP) have previously been described[3]. Mice were injected with $3 \times 10^{10}$ or $3 \times 10^{11}$ viral genomes (vg) of AAV-core-GFP 15 days before further experimental manipulation.

Third-generation, self-inactivating lentiviral vectors (LV.ET.mIL2.142T) that allow expression of mouse IL-2 exclusively in hepatocytes owing to the presence of a synthetic hepatocyte-specific promoter/enhancer as well as specific microRNA 142 target sequences that suppress expression in haematopoietic-lineage cells[39] were generated, produced and titrated as previously described[44]. In brief, the gene-synthesized mouse *Il2* cDNA was cloned into the previously described transfer vector pCCLsin.cPPT.ET.GFP.142T[44] by standard cloning techniques. Third-generation lentiviral vectors were produced by calcium phosphate transient transfection of 293T cells of the transfer vector, the packaging plasmid pMDLg/p.RRE, pCMV.REV, the vesicular stomatitis virus glycoprotein G (VSV-G) envelope plasmid pMD2.G and the pAdvantage plasmid (Promega), as previously described[44]. For integrase-defective lentiviral vector (IDLV) production, the pMDLg/p.RRE.D64Vint packaging with a mutant integrase was used instead of pMDLg/p.RRE, as described[45]. In brief, $9 \times 10^6$ 293T cells were seeded 24 h before transfection in 15-cm dishes. Two hours before transfection, culture medium was replaced with fresh medium. For each dish, a solution containing a mix of the selected transfer plasmid, the packaging plasmids pMDLg/pRRE and pCMV.REV, pMD2.G and the pAdvantage plasmid was prepared using 35, 12.5, 6.25, 9 and 15 μg of plasmid DNA, respectively. A 0.1× TE solution (10 mM Tris-HCl, 1 mM EDTA pH 8.0 in dH$_2$0) and water (1:2) was added to the DNA mix to 1,250 μl of final volume. The solution was left on a spinning wheel for 20–30 min, then 125 μl of 2.5 M CaCl$_2$ was added. Right before transfection, a precipitate was formed by adding 1,250 μl of 2× HBS (281 mM NaCl, 100 mM HEPES, 1.5

mM Na$_2$HPO$_4$, pH 7.12) while the solution was kept in agitation on a vortex. The precipitate was immediately added to the culture medium and left on cells for 14–16 h and after that the culture medium was changed. Supernatant was collected 30 h after medium change and passed through a 0.22-μm filter (Millipore). Filtered supernatant was transferred into sterile 25 × 89-mm polyallomer tubes (Beckman) and centrifuged at 20,000$g$ for 120 min at 20 °C (Beckman Optima XL-100K Ultracentrifuge). Vector pellet was dissolved in the appropriate volume of PBS to allow a 500× concentration. For lentiviral vector titration, $1 \times 10^5$ 293T cells were transduced with serial vector dilutions in the presence of polybrene (16 μg ml$^{-1}$). Genomic DNA (gDNA) was extracted 14 days after transduction. gDNA was extracted by using Maxwell 16 Cell DNA Purification Kit (Promega) according to manufacturer's instructions. Vector copies per diploid genome (vector copy number, VCN) were quantified by quantitative PCR (qPCR) starting from 100 ng of template gDNA using primers (HIV sense: 5′-TACTGACGCTCTCGCACC-3′; HIV antisense: 5′-TCTCGACGCAGGACTCG-3′) and a probe (FAM 5′-ATCTCTCTCCTTCTAGCCTC-3′) designed to amplify the primer binding site region of the lentiviral vector. Endogenous DNA amount was quantified by a primers/probe set designed to amplify the human telomerase gene (Telo sense: 5′-GGCACACGTGGCTTTTCG-3′; Telo antisense: 5′-GGTGAACCTCGTAAGTTTATGCAA-3′; Telo probe: VIC 5′-TCAGGACGTCGAGTGGACACGGTG-3′ TAMRA). Copies per genome were calculated by the formula = (ng LV/ng endogenous DNA) × (number of LV integrations in the standard curve), in which 'LV' denotes lentiviral vector. The standard curve was generated by using a CEM cell line stably carrying four vector integrants, which were previously determined by Southern blot and FISH analysis. All reactions were carried out in duplicate or triplicate in an ABI Prism 7900HT or Viia7 Real Time PCR thermal cycler (Applied Biosystems). Each qPCR run carried an internal control generated by using a CEM cell line stably carrying 1 vector integrant, which were previously determined by Southern blot and FISH analysis. Titre is expressed as transducing units$_{293T}$ (TU) per ml and calculated using the formula TU per ml = (VCN × 10$^5$ × 1/dilution factor). IDLV titre was determined on 293T cells 3 days after transduction using an ad hoc quantitative PCR, which selectively amplifies the reverse-transcribed vector genome (both integrated and non-integrated) discriminating it from plasmid carried over from the transient transfection (RT-LV; ΔU3 sense: 5′-TCACTCCCAACGAAGACAAGATC-3′, *gag* antisense: 5′-GAGTCCTGCGTCGAGAGAG-3′). Vector particles were measured by HIV-1 Gag p24 antigen immunocapture assay (Perkin Elmer) according to manufacturer's instructions. Vector infectivity was calculated as the ratio between titre and particles. Vector administration was carried out by tail vein injection in mice at $2.5 \times 10^8$–$10 \times 10^8$ TU per mouse, 7 days before T cell injection.

All infectious work was performed in designated BSL-2 or BSL-3 workspaces, in accordance with institutional guidelines.

**Naive T cell isolation, adoptive transfer and in vivo treatments.** CD8$^+$ T cells from the spleens of Cor93, Env28 and TCR-I transgenic mice were purified by negative immunomagnetic sorting (Miltenyi Biotec). Mice were adoptively transferred with $2 \times 10^4$–$5 \times 10^6$ CD8$^+$ T cells. In selected experiments, mice were splenectomized and treated with 200 μg of anti-CD62L monoclonal antibody (clone MEL-14, BioXcell) 48 h and 4 h before cell injection, respectively. Splenectomy was performed according to standard procedures[46]. In selected experiments, CD4$^+$ T cells were depleted by intravenously injecting 200 μg of anti-CD4 antibody (clone GK1.5, BioXcell) 3 days and 1 day before T cell transfer. In selected experiments, mice were treated with 200 μg of anti-PD-L1 (clone 10F.9G2, BioXcell) 1 day before and 1 day and 3 days after T cell transfer. In some experiments, T regulatory cells were depleted by intraperitoneally injecting 200 μg of purified anti-CD25 monoclonal antibodies (clone PC61, BioXcell) 8 days before T cell transfer. In selected experiments, wild-type or MUP-core mice were lethally irradiated and reconstituted with bone marrow from CD11c-DTR mice; dendritic cells were subsequently depleted by intraperitoneally injecting 25 ng g$^{-1}$ of diphtheria toxin (Sigma) 3 days and 1 day before T cell transfer. In indicated experiments, Kupffer cells were depleted by intravenous injection of clodronate-containing liposomes 2 days before T cell injection, as described[21]. IL-2–anti-IL-2 complexes (IL-2c) were prepared by mixing 1.5 μg of recombinant IL-2 (BioLegend) with 50 μg anti-IL-2 monoclonal antibody (clone S4B6-1, BioXcell) per mouse, as previously described[37]. Mice were injected with IL-2c intraperitoneally one day after T cell transfer, unless otherwise indicated.

**Cell isolation and flow cytometry.** Single-cell suspensions of livers, spleens, lymph nodes, bone marrow, lung and blood were generated as previously described[47,48]. Kupffer cell isolation was performed as previously described[3,49]. All flow cytometry stainings of surface-expressed and intracellular molecules were performed as previously described[3]. Cell viability was assessed by staining with Viobility 405/520 fixable dye (Miltenyi). Anitbodies used included: anti-CD3 (clone: 145-2C11, 562286, BD Biosciences), anti-CD11b (clone: M1/70, 101239), anti-CD19 (clone: 1D3, 562291 BD Biosciences), anti-CD25 (clone: PC61, 102015), anti-CD31 (clone: 390, 102427), anti-CD45 (clone: 30-F11, 564279 BD Biosciences),

anti-CD49b (clone: DX5, 562453 BD Biosciences), anti-CD64 (clone: X54-5/7.1, 139311), anti-F4/80 (clone: BM8, 123117), anti-I-A/I-E (clone: M5/114.15.2, 107622), anti-TIM4 (polyclonal, orb103599 Biorbyt), anti-CD69 (clone: H1.2F3, 104517), anti-CD45.1 (clone: A20, 110716), anti-IFNγ (clone: XMG1.2, 557735 BD Biosciences), anti-CD4 (clone: RM4-5, 553048 BD Biosciences), anti-CD11c (clone: N418, 117308), anti-I-Ab (clone: AF6-120.1, 116420), anti-PD-1 (clone: J43, 17-9985 eBioscience), anti-NK1.1 (clone: PK136, 108706), anti-NKp46 (clone: 29A1.4, 137623), anti-STAT5 pY694 (clone: 47, 560117 BD Biosciences), anti-FOXP3 (clone FJK-16s, 12-5773-80 eBioscience). All antibodies were purchased from BioLegend, unless otherwise indicated. Recombinant dimeric H-2L$^d$–Ig and H-2K$^b$–Ig fusion proteins (BD Biosciences) complexed with peptides derived from HBsAg (Env28-39) and HBcAg (Cor93-100), respectively, were prepared according to the manufacturer's instructions. Dimer staining was performed as previously described[47]. Flow cytometry staining for phosphorylated STAT5 was performed using Phosflow Perm Buffer III (558050, BD Bioscience), following the manufacturer's instructions. Flow cytometry staining for FOXP3 was performed using Foxp3/Transcription Factor Staining Buffer Set (00-5523-00, eBioscience), following the manufacturer's instructions.

All flow cytometry analyses were performed in FACS buffer containing PBS with 2 mM EDTA and 2% FBS on a FACS CANTO or LSRII (BD Biosciences) and analysed with FlowJo software (Treestar).

**Cell sorting.** Single-cell suspensions from spleens and livers were stained with Viobility 405/520 fixable dye (Miltenyi), with PB-conjugated anti-CD8α (clone 53-6.7) and PE-conjugated anti-CD45.1 antibodies. Live CD8$^+$ CD45.1$^+$ cells were sorted on a MoFlo Legacy (Beckman Coulter) cell sorter in a buffer containing PBS with 2% FBS. Cells were always at least 98% pure (data not shown).

**RNA purification and RNA-seq library preparation.** Total RNA was purified from 8,000–300,000 sorted cells with the ReliaPrep RNA Cell Miniprep System (Promega). Sequencing libraries were generated using the Smart-seq2 method[50]. In brief, 5 ng of RNA were retrotranscribed and cDNA was amplified using 15 PCR cycles and purified with AMPure XP beads (Beckman Coulter). After purification, the concentration was determined using Qubit 3.0 (Life Technologies) and the size distribution was assessed using Agilent 4200 TapeStation system. Then, the tagmentation reaction was performed starting from 0.5 ng of cDNA for 30 min at 55 °C and the enrichment PCR was carried out using 12 cycles. Libraries were then purified with AMPure XP beads, quantified using Qubit 3.0 and single-end sequenced (75 bp) on an Illumina NextSeq 500.

**RNA-seq data processing and analysis.** Reads were generated on a NextSeq 500 (Illumina) instrument following the manufacturer's recommendations. Single-end reads (75 bp) were aligned to the mm10 reference genome using STAR[51] aligner. featureCounts function from Rsubread package[52] was used to compute reads over RefSeq *Mus musculus* transcriptome, with option minMQS set to 255. Further analyses were performed with edgeR R package[53]. Pearson's correlation was computed for each couple of samples on log-transformed reads per kilobase per million (RPKM). Read counts were normalized with the trimmed mean of M-values (TMM) method[54] using calcNormFactors function and dispersion was estimated with the estimateDisp function. Differential expression across different conditions was evaluated fitting a negative binomial generalized linear model on the data set with glmQLFit function and then performing a quasi-likelihood $F$-test with glmQLFTest function. Batch information was included in the design as covariate.

**Differential gene expression analysis.** *Hepatic CD8$^+$ T cells from wild-type mice injected with rLCMV-core versus MUP-core mice.* Genes with an RPKM value higher than 1 in at least two samples in the datasets were retained. We first defined inducible genes, namely those genes with log$_2$-transformed fold change in RPKM (log$_2$FC$_{RPKM}$) > 2.5 and false discovery rate (FDR) < 0.01 relative to naive T cells in at least one condition or time point. For each comparison, only genes with an RPKM value higher than 1 in at least two samples in the comparison were selected. For each time point, induced genes were classified as expressed at higher levels in the WT + rLCMV-core condition setting FDR < 0.1 and log$_2$FC$_{RPKM}$ > 1.5 (WT + rLCMV-core vs MUP-core) as cut-off. Genes with an FDR < 0.1 and a log$_2$ FC$_{RPKM}$ < −1.5 in the WT + rLCMV-core vs MUP-core comparison were classified as expressed at higher levels in MUP-core. The remaining genes were defined as non-differentially expressed between WT + rLCMV-core and MUP-core.

*Hepatic or splenic CD8$^+$ T cells from wild-type mice injected with rLCMV-core versus Cor93 T$_N$ cells.* We first defined as expressed genes those having counts per million (CPM) > 1 in at least two samples in the dataset. Induced genes were defined using as cut-offs a log$_2$FC$_{RPKM}$ > 2.5 and FDR < 0.01 relative to naive T cells in at least one condition or time point. For each comparison, only genes with an RPKM value higher than 1 in at least two samples in the comparison were selected.

*Hepatic CD8$^+$ T cells from wild-type mice injected with rLCMV-core versus MUP-core mice with or without IL-2c treatment.* We first defined as expressed genes those having CPM >1 in at least two samples in the dataset. Induced genes were defined using as cut-offs a log$_2$FC$_{RPKM}$ > 2.5 and FDR < 0.01 relative to naive T cells in at least one condition. For each comparison, only genes with an RPKM value higher

than 1 in at least two samples in the comparison were selected. Induced genes were then classified as expressed at higher levels in the WT + rLCMV-core condition (hypo-expressed in MUP-core at day 5) setting log$_2$FC$_{RPKM}$ > 1.5 and FDR < 0.01 (WT + rLCMV-core vs MUP-core) as cut-off. Genes with a log$_2$FC$_{RPKM}$ < −1.5 and FDR < 0.1 in the WT + rLCMV-core vs MUP-core comparison were classified as expressed at higher levels (hyper-expressed in MUP-core at day 5) in MUP-core. We then classified genes hypo-expressed in MUP-core as rescued if they displayed log$_2$FC$_{RPKM}$ > 1 and FDR < 0.1 in the IL-2c-treated MUP-core versus MUP-core comparison. Conversely, genes hyper-expressed in MUP-core were defined as rescued if displaying log$_2$FC$_{RPKM}$ < −1 and FDR < 0.1 in the IL-2c-treated MUP-core versus MUP-core comparison. The remaining genes were classified as not rescued.

**GO analyses.** For each time point, we ranked expressed genes by decreasing order of log$_2$FC$_{RPKM}$ values in the WT + rLCMV-core versus MUP-core comparison. We then performed GSEA[42] on each of these ranked lists using the clusterProfiler R package[55] and the gene sets contained in the Biological Processes ontology from the org.Mm.eg.db database. GO categories with $q$ value < 0.1 were retained and aggregated using REVIGO[43] (similarity score = 0.7), yielding 143 seed GO categories showing enrichment in WT + rLCMV-core or in MUP-core in at least one time point.

**Gene expression analysis in published datasets.** RNA-seq and Sequence Read Archive (SRA) data were downloaded from the Gene Expression Omnibus (GEO) repository and converted to the FastQ format. Reads were then aligned against the whole *Mus musculus* mm10 genome build using STAR aligner (v.2.6.0a) with default options, generating BAM files. Read counts for all expressed genes (Ensembl annotation v.94; GENCODE M19) were obtained using featureCounts (Rsubread v.3.7). Features with <1 CPM were filtered out. The resulting count matrix was then normalized using the normalization factors generated by the upperquartile method[56] implemented in edgeR Bioconductor package. For Illumina BeadChip data, the normalized expression matrix was downloaded from the GEO repository. Genes with an expression level that corresponded to the 65th percentile of the distribution of the log$_2$(expression values) were considered to be expressed.

**ATAC-seq.** ATAC-seq was performed as previously described[57] with slight modifications. In brief, 8,000–50,000 cells per sample were sorted and centrifuged at 1,600 r.p.m. for 5 min. Then, the transposition reaction was performed using digitonin 1% (Promega), Tn5 transposase and TD Buffer (Illumina) for 45 min at 37 °C. Immediately after transposition, the reaction was stopped using a solution of 900 mM NaCl and 300 mM EDTA, 5% SDS and proteinase K (Sigma-Aldrich) for 30 min at 40 °C. Transposed DNA fragments were purified using AMPure XP beads (Beckman Coulter), barcoded with dual indexes (Illumina Nextera) and PCR amplified with KAPA HiFi PCR Kit (KAPA Biosystems). Then, the concentration of the library was determined using Qubit 3.0 (Life Technologies) and the size distribution was assessed using Agilent 4200 TapeStation system. Libraries were single-end sequenced (75 bp) on an Illumina NextSeq 500.

**ATAC-seq data processing and analysis.** Reads were generated on NextSeq 500 (Illumina) instrument following manufacturer's recommendations. Single-end reads (75 bp) were aligned to the mm10 reference genome using BWA[58] aligner. BAM files were processed using samtools[59] and BEDTools[60] suits: reads with a mapping quality lower than 15 or duplicated were discarded. Moreover, unassigned reads and reads mapped on chromosomes Y and M were removed. MACS2[61] callpeak function with parameters -g mm –nomodel –shift -100 –extsize 200 was used for peak calling. For each sample peaks with a $q$-value lower than 1e-10 were selected. Peaks from all samples that passed filter were then merged with mergeBed function form BEDTools, resulting in 72,884 regions. Reads counts were computed on this set of regions using coverageBed function from BEDTools. The set of 72,884 regions was annotated using ChIPpeakAnno R package[62]. Each region was associated to the gene with the closest transcription start site. Further analyses were performed with edgeR R package. Pearson's correlation was computed for each pair of samples on log-transformed CPM. As previously described for RNA-seq data, read counts were normalized with the TMM method using calcNormFactors function and dispersion was estimated with the estimateDisp function. Differences in peaks intensities across different conditions were evaluated fitting a negative binomial generalized linear model on the dataset with glmQLFit function and then performing a quasi-likelihood $F$-test with glmQLFTest function. Batch information was included in the design as covariate.

**Definition of induced and differentially induced ATAC-seq peaks.** We first defined inducible peaks, namely those regions with log$_2$-transformed fold change in CPM (log$_2$FC$_{CPM}$) > 2.5 and FDR < 0.001 relative to naive T cells in at least one condition or time point. For each time point, induced peaks were classified as induced at higher levels in the WT + rLCMV-core condition setting FDR < 0.1 and log$_2$FC$_{CPM}$ > 1.5 (WT + rLCMV-core versus MUP-core) as cut-offs. Peaks with an FDR < 0.1 and a log$_2$FC$_{CPM}$ < −1.5 in the WT + rLCMV-core versus MUP-core comparison were classified as induced at higher levels in MUP-core. The remaining

peaks were defined as non-differentially induced between WT + rLCMV-core and MUP-core.

**Motif enrichment analysis.** Enrichment analysis of known motifs was performed with HOMER[41] using findMotifsGenome.pl script. For each time point we ranked ATAC-seq peaks according to $\log_2FC_{CPM}$ values in the WT + rLCMV-core versus MUP-core comparison and selected the 200 regions showing highest or lowest $\log_2FC_{CPM}$ values. These sets of differentially induced regions were compared to a background composed by a set of 3,899 regions with unchanged intensities (FDR > 0.1 and absolute $\log_2FC_{CPM} < 0.5$) between both MUP-core and WT + rLCMV-core versus naive in all time points.

**Purification of viral nucleic acids from serum.** Twenty microlitres of serum was incubated for 2 h at 37 °C with 180 μl IsoHi buffer (150 mM NaCl, 0.5% NP40, 10 mM Tris pH 7.4), 5 mM CaCl$_2$, 5 mM MgCl2, 1 U DNaseI (Life Technologies), 5 U micrococcal nuclease (Life Technologies). The digestion was stopped by the addition of 20 mM EDTA pH 8.0 and viral nucleic acid purification performed with the QIAmp MiniElute Virus Spin Kit (Qiagen, 57704), according to the manufacturer's instructions.

**RT–qPCR.** Total RNA was extracted from frozen livers using ReliaPrep RNA Tissue Miniprep System (Promega), according to the manufacturer's instructions, as described[63], genomic DNA contamination was removed using Ambion TURBO DNA-freeTM DNase. 1 μg of total RNA was reverse transcribed with Superscript IV Vilo (Life Technologies) before qPCR analysis for mouse *Il2* (TaqMan Mm00434256, Life Technologies), *Ifng* (TaqMan Mm01168134, Life Technologies), HBV core (forward TACCGCCTCAGCTCTGTATC, reverse CTTCCAAATTAACACCCACCC, probe TCACCTCACCATACTGCACTCAGGCAA). Reactions were run and analysed on Quant Studio 5 instrument (Life Technologies). For viraemia quantification, a standard curve was drawn using plasmid DNA. All experiments were performed in triplicate and normalized to the reference gene *Gapdh*.

**Western blot analysis.** Western blot analysis on frozen liver homogenates or on Kupffer cells was performed exactly as previously described[64]. Primary antibodies include anti-STAT5 and anti-pSTAT5 (Tyr694) (rabbit; Cell Signaling 8215), anti-HBcAg (polyclonal, Dako), β-actin (polyclonal; Abcam ab228001) and H3 (polyclonal; abcam ab1791). Secondary antibodies include horseradish peroxidase-conjugated goat anti-rabbit IgG (Jackson ImmunoResearch). Reactive proteins were visualized using a Clarity Western ECL substrate kit (Bio-Rad), and exposure was performed using UVItec (Cambridge MINI HD, Eppendorf). Images were acquired by NineAlliance software. Band quantification was performed with ImageJ software on 16-bit images and normalized on the matching housekeeping protein as a loading control. Each lane corresponds to a different mouse.

**Southern blot analysis.** Southern blot analysis on total DNA isolated from frozen livers (left lobe) was performed exactly as previously described[63].

**Confocal immunofluorescence histology and histochemistry.** Confocal microscopy analysis of livers was performed as previously described[3]. The following primary antibodies were used for staining: anti-F4/80 (BM8, Invitrogen), anti-cytokeratin 7 (EPR17078, Abcam), anti-LYVE-1 (NB600-1008, Novus Biological), anti-HBcAg (polyclonal, Dako). The following secondary antibodies were used for staining: Alexa Fluor 488-, Alexa Fluor 514-, Alexa Fluor 568-, or Alexa Fluor 647-conjugated anti-rabbit or anti-rat IgG (Life Technologies). Images were acquired on an inverted Leica microscope (TCS STED CW SP8, Leica Microsystems) with a motorized stage for tiled imaging. To minimize fluorophore spectral spillover, we used the Leica sequential laser excitation and detection modality. The bleed-through among sequential fluorophore emission was removed applying simple compensation correction algorithms to the acquired images. The semiautomatic surface-rendering module in Imaris (Bitplane) was used to create 3D volumetric surface objects corresponding either to individual cells or to the liver vascular system. Signal thresholds were determined using the Imaris Surface Creation module, which provides automatic threshold. T cells were tracked manually for single cell distance from the centre of each bile duct (CK7$^+$) using Fiji.

For H&E and HBcAg immunohistochemistry, livers were perfused with PBS, collected in Zn-formalin and transferred into 70% ethanol 24 h later. Tissue was then processed, embedded in paraffin and stained as previously described[3]. Bright-field images were acquired through an Aperio Scanscope System CS2 microscope and an ImageScope program (Leica Biosystem) following the manufacturer's instructions.

**Intravital multiphoton microscopy.** Liver intravital multiphoton microscopy was performed as previously described[3,65]. Liver sinusoids were visualized by injecting nontargeted Quantum Dots 655 (Invitrogen) intravenously during image acquisition. Images were acquired with a LaVision BioTec TriMScope II coupled to a Nikon Ti-U inverted microscope enclosed in a custom-built environmental chamber (Life Imaging Services) that was maintained at 37–38 °C with heated air. Continuous body temperature monitoring through a rectal probe was performed to ensure that a narrow range of 37–38 °C was maintained at all times. Fluorescence excitation was provided by two tuneable femtosecond (fs)- pulsed Ti:Sa lasers (680–1,080 nm, 120 fs pulse-width, 80 MHz repetition rate, Ultra II, Coherent), an Optical Parametric Oscillator (1,000–1,600 nm, 200 fs pulse-width, 80 MHz repetition rate, Chameleon Compact OPO, Coherent). The setup includes four non-descanned photomultiplier tubes (Hamamatsu H7422-40 GaAsP High Sensitivity PMTs and Hamamatsu H7422-50 GaAsP High Sensitivity red-extended PMT from Hamamatsu Photonics K.K.), a 25×, 1.05 NA, 2 mm working distance, water-immersion multiphoton objective (Olympus). For 4D analysis of cell migration, stacks of 7–15 square *xy* sections (512 × 512 pixel) sampled with 4 μm *z* spacing were acquired every 5–32 s for up to 2 h, to provide image volumes that were 40 μm in depth and with an *xy* field of view variable between 100 × 100 μm$^2$ and 450 × 450 μm$^2$. Sequences of image stacks were transformed into volume-rendered, 4D time-lapse movies with Imaris (Bitplane). The 3D positions of the cell centroids were segmented by semi-automated cell tracking algorithm of Imaris. The semiautomatic surface-rendering module in Imaris (Bitplane) was used to create 3D volumetric surface objects corresponding either to individual cells or to the liver vascular system. Signal thresholds were determined using the Imaris Surface Creation module, which provides automatic threshold.

**Biochemical analyses.** The extent of hepatocellular injury was monitored by measuring sALT activity at multiple time points after treatment, as previously described[3].

**Patients and study approval.** A total of 29 patients with chronic HBV infection (HBsAg$^+$) were included. The patients were subdivided into the disease categories immune-tolerant and immune-active, on the basis of their clinical history (Supplementary Table 10). In brief, the 13 immune-tolerant patients had no history of hepatitis (normal ALT) and were all positive for HBeAg. The 16 immune-active patients (4 HBeAg$^+$ and 12 HBeAg$^-$) have or had previously signs of hepatic inflammation (ALT > 40 infectious units l$^{-1}$), 6 of them are currently or were previously treated with nucleoside analogues. Supplementary Table 10 summarizes the available clinical and virological parameters. Blood donors were recruited from the viral hepatitis clinic at The Royal London Hospital. Written informed consent was obtained from all subjects. The study was conducted in accordance with the Declaration of Helsinki and approved by the Barts and the London NHS Trust local ethics review board and the NRES Committee London–Research Ethics Committee (reference 10/H0715/39) and by the Singapore National Healthcare Group ethical review board (DSRB 2008/00293).

**Clinical and virological parameters.** On recruitment to the study, viral serology and HBV DNA levels were tested. HBsAg, HBeAg and anti-HBe levels were measured with a chemiluminescent microparticle immunoassay (CMIA; Architect Assay, Abbott Diagnostics). HBV DNA levels in serum were quantified by real-time PCR (COBAS AmpliPrep/COBAS TaqMan HBV test v.2.0; Roche Molecular Diagnostics) and HBV genotyping was performed by restriction fragment length polymorphism analysis of a pre-S amplicon, as previously described[66].

**HBV peptide library.** Three libraries of 311–313 15-mer peptides overlapping by 10 amino acids were used to identify HBV-specific T cells. The peptides covered the entire sequence of HBV genotypes B, C and D (GenBank AF121243, AF 112063 and AF 21241, respectively) and were purchased from Mimotopes. The purity of the peptides was above 80%, and their composition was confirmed by mass spectrometry analysis. Peptides were pooled as previously described[67]. The peptide libraries were matched to the HBV genotype of each patient as indicated in Supplementary Table 10. For patients infected with HBV genotype A or E, the peptide library of genotype D was used.

**Peripheral blood mononuclear cell isolation and T cell culture.** Peripheral blood mononuclear cells were isolated from peripheral blood by Ficoll gradient and cryopreserved. Cells were thawed, and T cell lines were generated as follows: 20% of peripheral blood mononuclear cells were pulsed with 10 μg ml$^{-1}$ of the overlapping HBV peptides for 1 h at 37 °C, subsequently washed, and cocultured with the remaining cells in AIM-V medium (Gibco; Thermo Fisher Scientific) supplemented with 2% AB human serum (Gibco; Thermo Fisher Scientific). T cell lines were cultured for 10 days, with or without the presence of 20 U ml$^{-1}$ of recombinant IL-2 (R&D Systems).

**ELISpot assays.** ELISpot assays for the detection of IFNγ-producing cells were performed on in vitro expanded T cell lines using HBV peptides pooled into the following mixtures: X, core, Env1, Env2, Pol1, Pol2, Pol3 and Pol4. T cell lines were incubated overnight at 37 °C with pools of HBV peptides (1 μg ml$^{-1}$), in which final DMSO concentrations did not exceed 0.2%. Medium was supplemented as before with or without 20 U ml$^{-1}$ of recombinant IL-2. IFNγ ELISpot assays (Millipore) were performed as previously described[67].

**Statistical analyses.** Results are expressed as mean ± s.e.m. All statistical analyses were performed in Prism (GraphPad Software), and details are provided in the figure legends.

**Reporting summary.** Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

## Data and availability

The RNA-seq and ATAC-seq data on sorted hepatic CD8$^+$ T cells have been deposited in the ArrayExpress database under the accession codes E-MTAB-7462 and E-MTAB-7461, respectively. All other data are available in the main text or the supplementary materials.

44. Cantore, A. et al. Liver-directed lentiviral gene therapy in a dog model of hemophilia B. *Sci. Transl. Med.* **7**, 277ra28–277ra28 (2015).
45. Mátrai, J. et al. Hepatocyte-targeted expression by integrase-defective lentiviral vectors induces antigen-specific tolerance in mice with low genotoxic risk. *Hepatology* **53**, 1696–1707 (2011).
46. Reeves, J. P., Reeves, P. A. & Chin, L. T. Survival surgery: removal of the spleen or thymus. *Curr. Protoc. Immunol.* Chapter 1, Unit 1.10 (2001).
47. Iannacone, M. et al. Platelets mediate cytotoxic T lymphocyte-induced liver damage. *Nat. Med.* **11**, 1167–1169 (2005).
48. Tonti, E. et al. Bisphosphonates target B cells to enhance humoral immune responses. *Cell Reports* **5**, 323–330 (2013).
49. Li, P.-Z., Li, J.-Z., Li, M., Gong, J.-P. & He, K. An efficient method to isolate and culture mouse Kupffer cells. *Immunol. Lett.* **158**, 52–56 (2014).
50. Picelli, S. et al. Full-length RNA-seq from single cells using Smart-seq2. *Nat. Protocols* **9**, 171–181 (2014).
51. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
52. Liao, Y., Smyth, G. K. & Shi, W. The Subread aligner: fast, accurate and scalable read mapping by seed-and-vote. *Nucleic Acids Res.* **41**, e108–e108 (2013).
53. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2010).
54. Robinson, M. D. & Oshlack, A. A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol.* **11**, R25 (2010).
55. Yu, G., Wang, L.-G., Han, Y. & He, Q.-Y. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS* **16**, 284–287 (2012).
56. Bullard, J. H., Purdom, E., Hansen, K. D. & Dudoit, S. Evaluation of statistical methods for normalization and differential expression in mRNA-Seq experiments. *BMC Bioinformatics* **11**, 94 (2010).
57. Buenrostro, J. D., Wu, B., Chang, H. Y. & Greenleaf, W. J. ATAC-seq: a method for assaying chromatin accessibility genome-wide. *Curr. Protoc. Mol. Biol.* **109**, 21.29.1–9 (2015).
58. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
59. Li, H. et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
60. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).
61. Zhang, Y. et al. Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9**, R137 (2008).
62. Zhu, L. J. et al. ChIPpeakAnno: a Bioconductor package to annotate ChIP-seq and ChIP-chip data. *BMC Bioinformatics* **11**, 237 (2010).
63. Fioravanti, J. et al. Effector CD8$^+$ T cell-derived interleukin-10 enhances acute liver immunopathology. *J. Hepatol.* **67**, 543–548 (2017).
64. Zordan, P. et al. Tuberous sclerosis complex-associated CNS abnormalities depend on hyperactivation of mTORC1 and Akt. *J. Clin. Invest.* **128**, 1688–1706 (2018).
65. Benechet, A. P., Ganzer, L. & Iannacone, M. Intravital microscopy analysis of hepatic T cell dynamics. *Methods Mol. Biol.* **1514**, 49–61 (2017).
66. Lindh, M., Gonzalez, J. E., Norkrans, G. & Horal, P. Genotyping of hepatitis B virus by restriction pattern analysis of a pre-S amplicon. *J. Virol. Methods* **72**, 163–174 (1998).
67. Tan, A. T. et al. Host ethnicity and virus genotype shape the hepatitis B virus-specific T-cell repertoire. *J. Virol.* **82**, 10986–10997 (2008).
68. Barbier, L. et al. Two lymph nodes draining the mouse liver are the preferential site of DC migration and T cell activation. *J. Hepatol.* **57**, 352–358 (2012).
69. Thierry, G. R. et al. The conduit system exports locally secreted IgM from lymph nodes. *J. Exp. Med.* **245**, jem.20180344 (2018).

**Extended Data Fig. 1** | See next page for caption.

**Extended Data Fig. 1 | Naive CD8$^+$ T cells that recognize hepatocellular antigen are activated locally and expand, but do not develop effector function. a**, Schematic of the experimental setup. Five million Env28 $T_N$ cells were transferred into C57BL/6 × BALB/c $F_1$ (WT) or HBV replication-competent transgenic (HBV Tg, C57BL/6 × BALB/c $F_1$) recipients. Livers were collected and analysed 5 days after Env28 $T_N$ cell transfer and sera from the same mice were collected daily from day 0 to 5 after transfer. **b, c**, Absolute numbers of total (**b**) and IFN$\gamma$-producing (**c**) Env28 T cells in the livers of the indicated mice. **d**, ALT levels detected in the sera of the indicated mice at the indicated time points. $n = 4$. **e**, Schematic of the experimental setup. Five million Cor93 $T_N$ cells were transferred into C57BL/6 (WT) or MUP-core recipients. Mice were splenectomized and treated with anti-CD62L antibody 48 h or 4 h before cell transfer, respectively. Untreated wild-type mice that received $5 × 10^6$ Cor93 $T_N$ cells were used as controls. Where indicated, mice were injected with $2.5 × 10^5$ infectious units of non-replicating rLCMV-core 4 h before Cor93 $T_N$ cell transfer. Liver-draining lymph nodes[68] (dLN) and non-draining inguinal lymph nodes (ndLN) were collected at 4 h and 1 day after transfer. **f**, Representative flow cytometry plots 4 h after Cor93 $T_N$ cells transfer. Numbers indicate the percentage of cells within the indicated gate. **g, h**, Quantification of the absolute numbers of cells recovered from the ndLN (**g**) and dLN (**h**) of the indicated mice 4 h and 1 day (d1) after Cor93 $T_N$ cell transfer. $n = 3$. **i**, Confocal immunofluorescence micrographs of liver sections from wild-type mice, wild-type mice transduced with rLCMV-core, MUP-core mice, and R26-ZsGreen mice injected with $2.5 × 10^5$ infectious units of non-replicating rLCMV-cre. Scale bars, 100 μm. Note that, because HBV core protein did not accumulate at detectable levels in Kupffer cells and hepatic dendritic cells after rLCMV-core injection, we confirmed the tropism of this vector by injecting rLCMV-cre into R26-ZsGreen mice, which express the fluorescent protein ZsGreen after Cre-mediated recombination. **j**, MFI of CD69 expression on Cor93 T cells in the liver, blood, lung and bone marrow of the indicated mice 4 h after Cor93 $T_N$ cell transfer. $n = 4$. Data are mean ± s.e.m. and representative of at least three independent experiments. \*\*\*$P < 0.001$, two-tailed $t$-test (**b, c**) or one-way ANOVA with Bonferroni post-test (**g, h, j**). Mouse drawings were adapted from ref.[69].
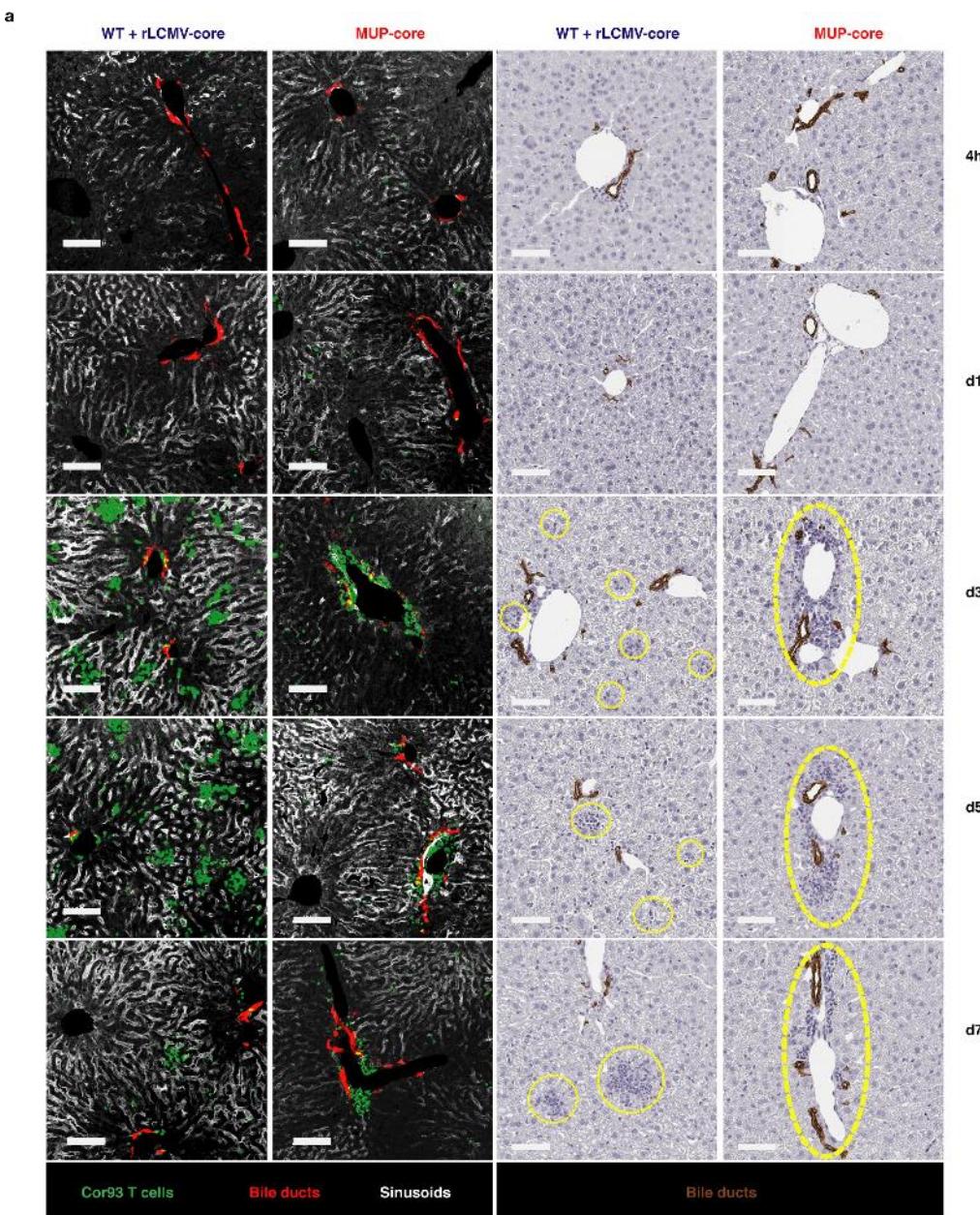
**Extended Data Fig. 2 | Spatiotemporal dynamics of naive CD8$^+$ T cells after intrahepatic priming.** Five million fluorescent Cor93 T$_N$ cells were transferred into MUP-core mice or wild-type mice transduced with rLCMV-core. Mice were splenectomized and treated with anti-CD62L antibody 48 h or 4 h before Cor93 T$_N$ transfer cell, respectively. **a**, Left, confocal immunofluorescence micrographs of liver sections from the indicated mice at the indicated time points after Cor93 T$_N$ cell transfer, showing the distribution of Cor93 T cells (green) relative to portal tracts (highlighted by anti-cytokeratin 7 (CK-7)-antibody-mediated staining of bile ducts in red). Sinusoids are highlighted by anti-LYVE-1$^+$ antibodies (white). Scale bars, 100 μm. Right, immunohistochemical micrographs of liver sections from the indicated mice at the indicated time points after Cor93 T$_N$ cell transfer, showing the distribution of leukocyte infiltrates relative to portal tracts (highlighted by anti-CK-7-antibody-mediated staining of bile ducts in brown). Scale bars, 100 μm. **b**, Distribution of the distances of each Cor93 T cell from the centre of the closest portal triad at the indicated time points. $n = 3$ mice. Data are representative of at least three independent experiments.
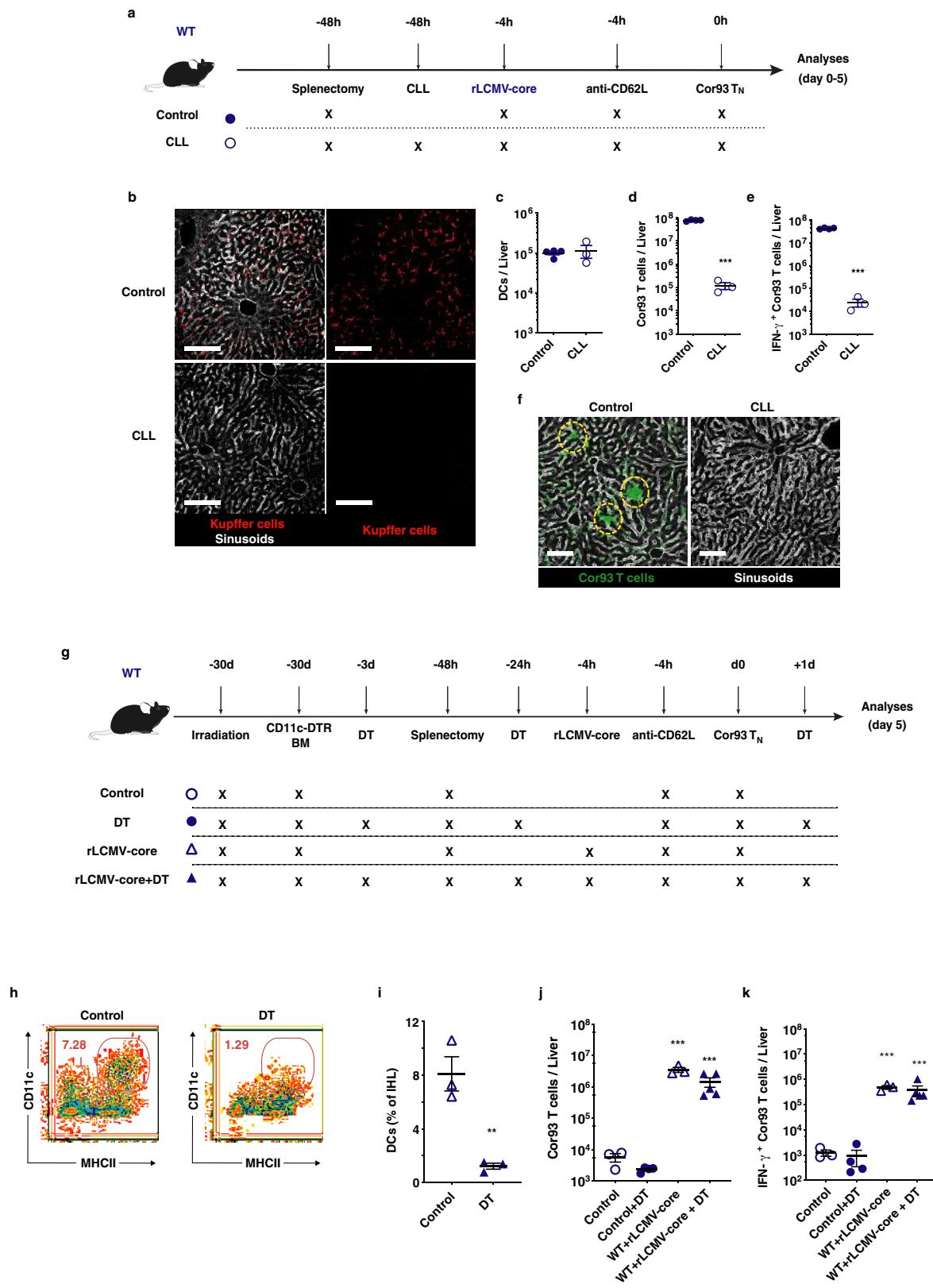
**Extended Data Fig. 3 |** See next page for caption.

**Extended Data Fig. 3 | Kupffer cells, but not dendritic cells, promote CD8$^+$ T cell effector differentiation after rLCMV injection.**
**a**, Schematic of the experimental setup. Five million Cor93 T$_N$ cells were transferred into C57BL/6 (WT) recipients. Mice were splenectomized and treated with anti-CD62L antibodies 48 h or 4 h before cell transfer, and injected with $2.5 \times 10^5$ infectious units of non-replicating rLCMV-core 4 h before Cor93 T$_N$ cell transfer. Where indicated, mice were treated with clodronate liposomes (CLL) 48 h before Cor93 T$_N$ cell transfer.
**b**, Confocal microscopy of liver sections from control mice (top) and clodronate liposome-treated mice (bottom) Kupffer cells are depicted in red in all panels, and sinusoids are depicted in grey only in the left panels. Scale bars, 100 μm. **c**, Absolute numbers of CD11c$^+$MHC-II$^{high}$ dendritic cells (DCs) in the livers of the indicated mice. **d**, **e**, Absolute numbers of total (**d**) and of IFNγ-producing (**e**) Cor93 T cells in the livers of the indicated mice 5 days after Cor93 T$_N$ cell transfer. $n = 4$ mice (control) and 3 mice (CLL). **f**, Confocal immunofluorescence micrographs of liver sections from the indicated mice 5 days after Cor93 T$_N$ cell transfer. Scale bars, 100 μm. **g**, Schematic of the experimental setup. Wild-type mice

were lethally irradiated and reconstituted with CD11c-DTR bone marrow (BM). One million Cor93 T$_N$ cells were transferred into recipients. Mice were injected with $2.5 \times 10^5$ infectious units of non-replicating rLCMV-core 4 h before Cor93 T$_N$ cell transfer. Indicated mice were treated with 400 ng of diphtheria toxin (DT) 3 days before, 1 day before and 1 day after T cell transfer. Livers were collected and analysed 5 days after Cor93 T$_N$ cell transfer. **h**, Representative flow cytometry plots in the liver of control (left) or diphtheria-toxin-treated (right) mice. **i**, CD11c$^+$MHC-II$^+$ dendritic cells (expressed as percentage of the total intrahepatic leukocyte population, IHL) in the livers of the indicated mice. $n = 3$. **j**, **k**, Absolute numbers of total (**j**) and IFNγ-producing (**k**) Cor93 T cells in the livers of the indicated mice 5 days after Cor93 T$_N$ cell transfer. $n = 3$ (control and WT + rLCMV-core), 4 (control + DT) and 5 (WT + rLCMV-core + DT). Data are mean ± s.e.m. and representative of three independent experiments. **$P < 0.01$, ***$P < 0.001$, two-tailed $t$-test (**d**, **e**, **i**) or one-way ANOVA with Bonferroni post-test (**i**–**k**). Mouse drawings were adapted from ref. [69].
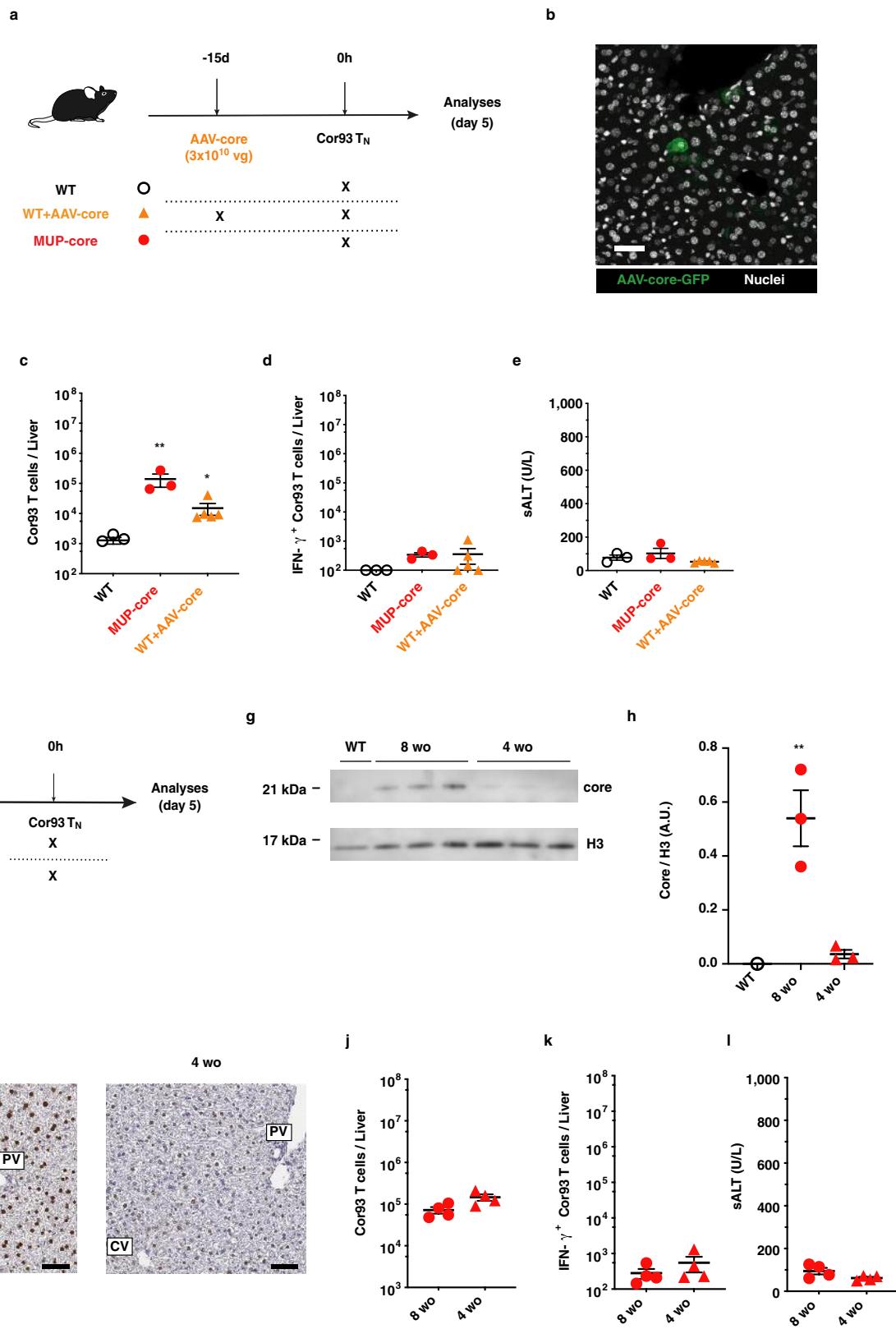
**Extended Data Fig. 4 |** See next page for caption.

**Extended Data Fig. 4 | A strong reduction in the levels of hepatocellular core antigen expression is per se not sufficient to induce effector differentiation. a**, Schematic of the experimental setup. One million Cor93 $T_N$ cells were transferred into C57BL/6 (WT) or MUP-core recipients. Indicated wild-type mice were injected with $3 \times 10^{10}$ viral genomes of AAV-core 15 days before Cor93 $T_N$ cell transfer. Livers were collected and analysed 5 days after Cor93 $T_N$ cell transfer. **b**, Representative confocal immunofluorescence micrographs of a liver section from an AAV-core-injected mouse 15 days after virus injection. Transduced hepatocytes are depicted in green and nuclei in grey. Scale bar, 50 μm. $n = 3$ mice. **c–e**, Absolute numbers of total (**c**) and IFNγ-producing (**d**) Cor93 T cells in the livers of the indicated mice 5 days after Cor93 $T_N$ cell transfer. **e**, ALT levels detected in the sera of the indicated mice. $n = 3$ (WT and MUP-core) and 5 (AAV-core). **f**, Schematic of the experimental setup. One million Cor93 $T_N$ cells were transferred into 8- or 4-week-old (wo) MUP-core mice. Livers were collected and analysed 5 days after Cor93 $T_N$ cell transfer. **g**, Expression of HBV core antigen (HBcAg) in the livers of the indicated mice was analysed by western blotting. **h**, Quantification of the western blot shown in **g**. Core expression, normalized to the housekeeping nuclear protein H3, is expressed as arbitrary units (A.U.). $n = 1$ (WT) and 3 (MUP-core 8wo and MUP-core 4wo). **i**, Immunohistochemical micrographs of liver sections from the indicated mice, showing core antigen expression (brown). Scale bars, 50 μm. CV, central vein; PV, portal vein. $n = 3$ mice. **j**, **k**, Absolute numbers of total (**j**) and of IFNγ-producing (**k**) Cor93 T cells in the livers of the indicated mice 5 days after Cor93 $T_N$ cell transfer. $n = 4$ mice. **l**, ALT levels detected in the sera of the indicated mice. $n = 4$. Data are mean ± s.e.m. and representative of two independent experiments. $*P < 0.05$, $**P < 0.01$, one-way ANOVA with Bonferroni post-test (**c–e**) or two-tailed $t$-test (**h**, **j–l**). Mouse drawings were adapted from ref. [69].
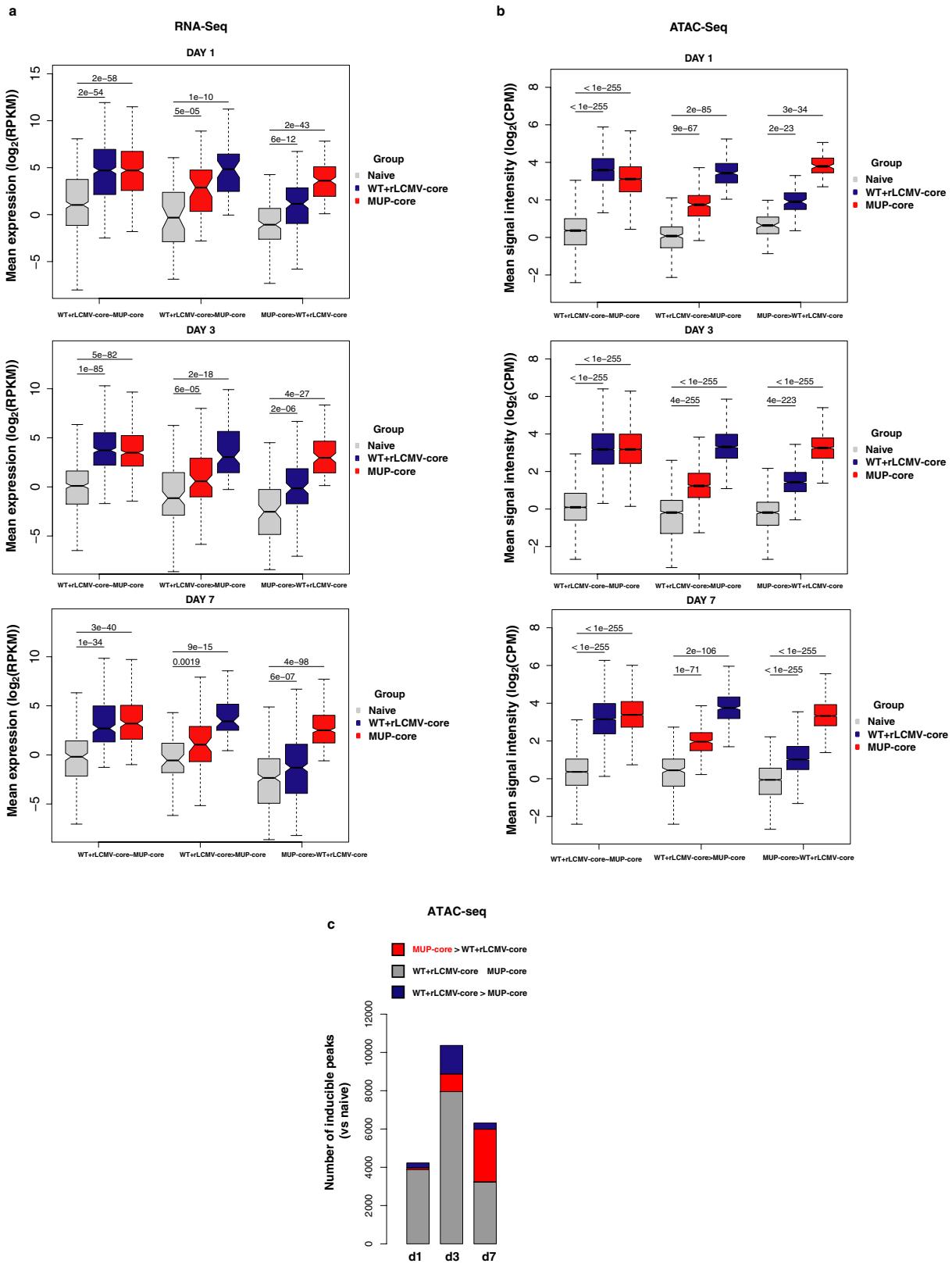
**Extended Data Fig. 5 |** See next page for caption.

**Extended Data Fig. 5 | Genomic landscape of naive CD8$^+$ T cells undergoing intrahepatic priming. a**, Box plots showing expression levels (log$_2$(RPKM)) in the indicated experimental condition of genes belonging to the categories described in Fig. 2a. Box plots are as in Fig. 3f. Naive ($n = 2$), WT + rLCMV-core ($n = 3$), MUP-core (day1 and 3, $n = 2$; day 7, $n = 3$). **b**, Box plots showing ATAC-seq signal intensity (log$_2$(CPM)) in the indicated experimental condition of peaks belonging to the categories described in Fig. 2c. Box plots are as in Fig. 3f. Naive ($n = 2$), WT + rLCMV-core (day 1 and 7, $n = 2$; day 3, $n = 3$), MUP-core (day 1 and 3, $n = 2$; day 7, $n = 3$). $P$ values in **a** and **b** were determined by two-sided Mann–Whitney $U$-test. **c**, Bar plot showing the number of inducible ATAC-seq peaks (logFC$_{CPM} > 2.5$, FDR $< 0.001$ versus Cor93 T$_N$) in the indicated conditions. ATAC-seq peaks with higher intensity signal in Cor93 T cells from WT + rLCMV-core (logFC$_{CPM} > 1.5$, FDR $< 0.1$) or from MUP-core mice (logFC$_{CPM} < -1.5$, FDR $< 0.1$) are shown in blue and red, respectively. Differences in peak signal intensities were evaluated fitting a negative binomial generalized linear model on the dataset and then performing a quasi-likelihood $F$-test. The Benjamini–Hochberg procedure was applied to correct for multiple tests. Naive ($n = 2$), WT + rLCMV-core (day 1 and 7, $n = 2$; day 3, $n = 3$), MUP-core (day 1 and 3, $n = 2$; day 7, $n = 3$).
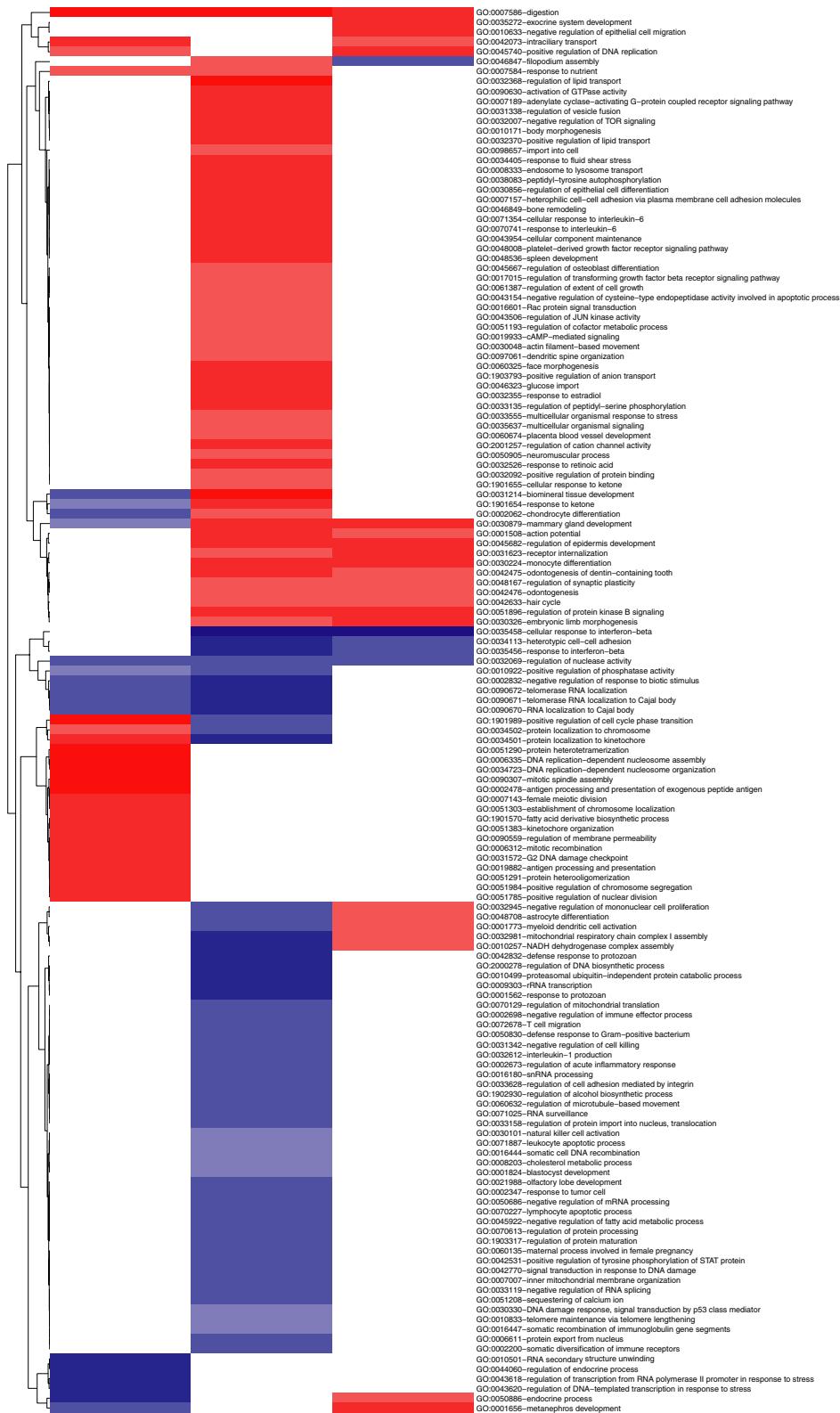
**Extended Data Fig. 6 | GO analysis of intrahepatically primed CD8+ T cells.** Heat map showing the NES value associated to the seed GO categories (identified by REVIGO) found enriched in the indicated time points by GSEA. Colour legends indicate NES, with positive values (in blue) reflecting enrichment of GO categories in hepatic CD8+ T cells isolated from wild-type mice injected with rLCMV-core, and negative values (in red) reflecting enrichment of GO categories in hepatic CD8+ T cells isolated from MUP-core mice.

**a** Expression of WT+rLCMV-core genes (day 7) in Scott-Browne et al.

**b** Expression of MUP-core genes (day 7) in Scott-Browne et al.

**c** Expression of WT+rLCMV-core genes (day 7) in Pauken et al.

**d** Expression of MUP-core genes (day 7) in Pauken et al.

**e** Expression of WT+rLCMV-core genes (day 7) in Schietinger et al.

**f** Expression of MUP-core genes (day 7) in Schietinger et al.

Gene set: CD8+ T cell exhausted genes (Scott-Browne)

**g** Day 1

NES -1.19
*p value* = 0.19

WT+rLCMV-core          MUP-core

**h** Day 3

NES -1.47
*p value* = 0.03

WT+rLCMV-core          MUP-core

**i** Day 7

NES -2.15
*p value* < 0.01

WT+rLCMV-core          MUP-core

**Extended Data Fig. 7** | See next page for caption.

**Extended Data Fig. 7 | Although priming by hepatocytes initiates a unique dysfunctional program, hepatocellular antigen persistence may gradually trigger an additional exhaustion signature. a–d**, Left, number of top 100 genes from Cor93 T cells recovered from the livers of wild-type mice transduced with rLCMV-core (**a**, **c**) or of MUP-core (**b**, **d**) mice reaching $\log_2(\text{RPKM}) > 1$ in the indicated conditions in RNA-seq data from splenic LCMV-specific effector or exhausted CD8[+] T cells[30] (**a**, **b**) or splenic LCMV-specific exhausted CD8[+] T cells[31] (**c**, **d**). Right, box plots showing the expression levels of top 100 genes from Cor93 T cells recovered from livers of WT + rLCMV-core mice (**a**, **c**) or MUP-core mice (**b**, **d**) in the indicated conditions in RNA-seq data from splenic LCMV-specific effector or exhausted CD8[+] T cells[30] (**a**, **b**) or splenic LCMV-specific exhausted CD8[+] T cells[31] (**c**, **d**). Naive ($n = 2$), effector ($n = 2$), exhausted ($n = 2$). **e**, **f**, Left, number of top 100 genes in Cor93 T cells isolated from livers of wild-type + rLCMV-core mice (**e**) or from MUP-core mice (**f**) expressed ($\log_2(\text{normalized data}) > 65$th percentile of the full distribution) in the indicated conditions in microarray data from

tolerant self-antigen-specific CD8[+] T cells[32]. Right, box plots showing the expression levels of genes retrieved in the dataset among the top 100 genes in Cor93 T cells isolated from the livers of wild-type + rLCMV-core (**e**) or MUP-core (**f**) mice in the indicated conditions in microarray data from tolerant self-antigen-specific CD8[+] T cells[32]. Only genes for which microarray probes were retrieved were kept for these analyses. Naive ($n = 3$), tolerant ($n = 3$). *$P < 0.05$, **$P < 0.01$, ***$P < 0.001$, two-tailed Wilcoxon rank-sum test. All box plots are represented as in Fig. 3f, and dots represent the expression distribution of the set of 100 genes. **g–i**, Enrichment plot showing the results of a GSEAPreanked analysis (Kolmogorow–Smirnov statistics) performed on genes expressed in CD8[+] T cells from wild-type + rLCMV-core or MUP-core mice (gene lists ranked by $\log(\text{FC}_{\text{RPKM}})$) and using as gene set a curated list of genes induced in exhausted CD8[+] T cells ($n = 2$) but not in effector CD8[+] T cells ($n = 2$) as compared to naive cells ($n = 2$)[30]. NES and $P$ values are reported for each time point.
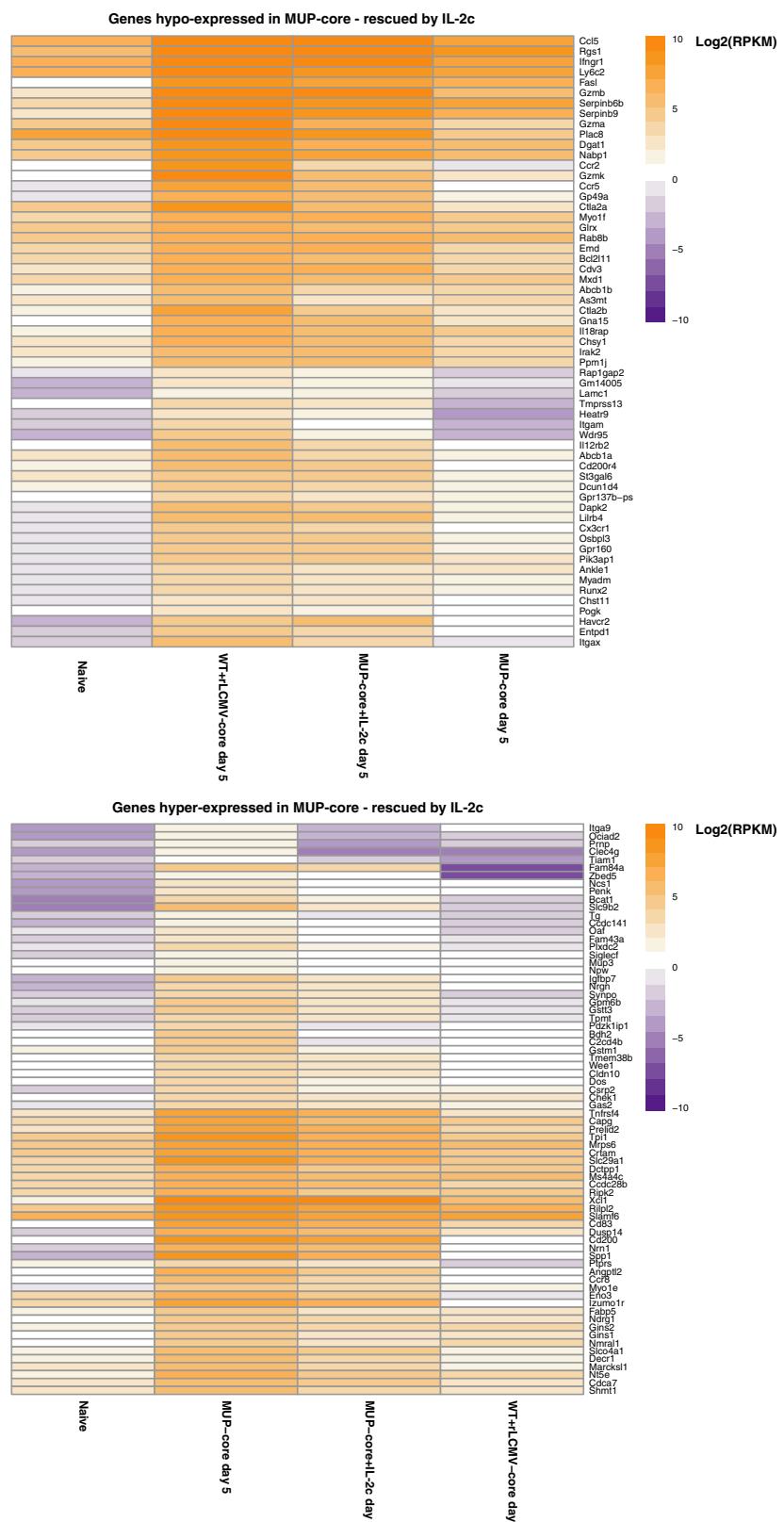
**Extended Data Fig. 8 | IL-2c substantially rescues the transcriptional program of dysfunctional CD8⁺ T cells.** Heat map showing expression values (log₂(RPKM)) of genes hypo-expressed (top) or hyper-expressed (bottom) in Cor93 CD8⁺ T cells from livers of MUP-core mice at day 5, which are rescued by treatment with IL-2c.
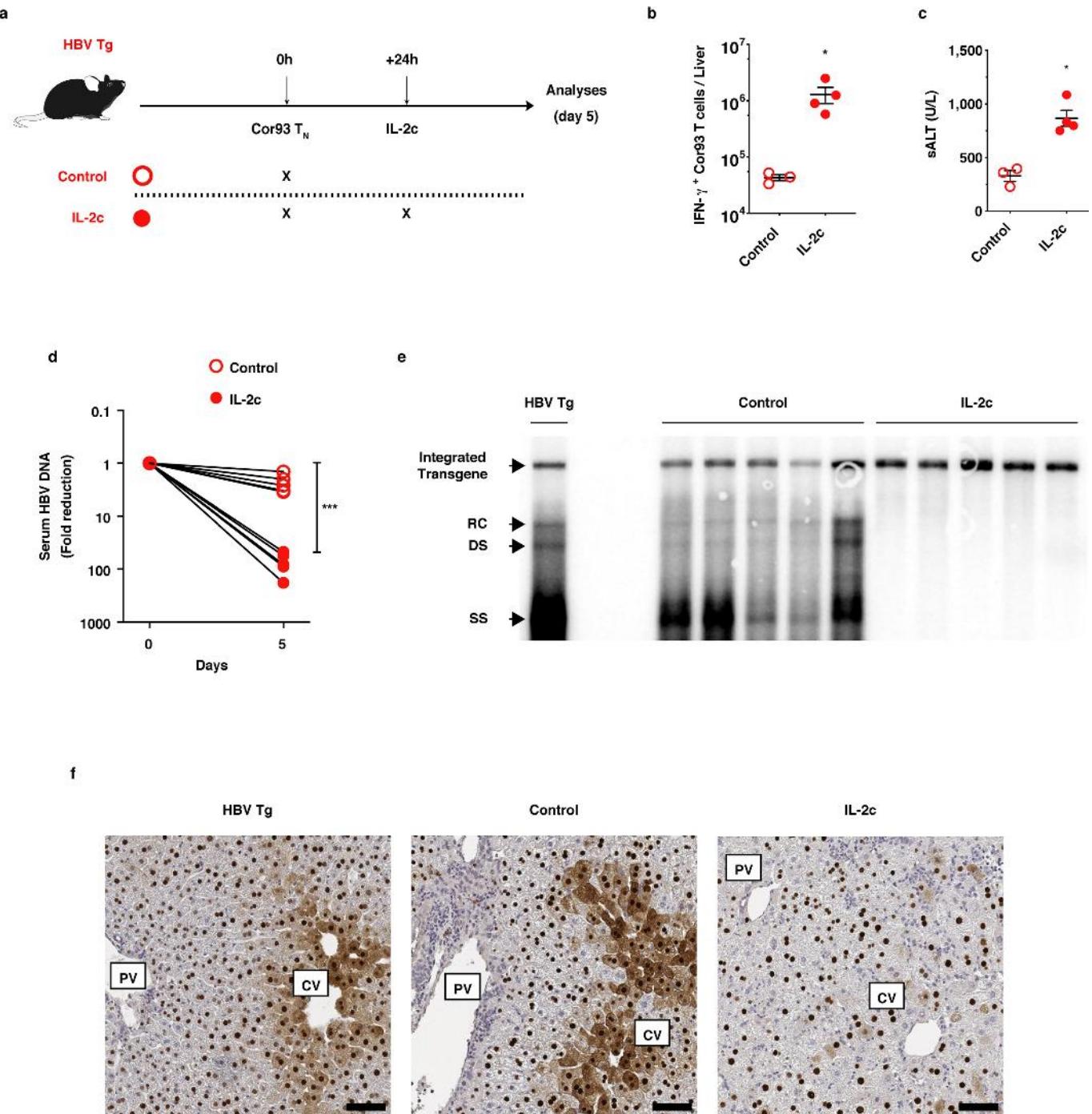
**Extended Data Fig. 9 | Therapeutic restoration of intrahepatically primed, dysfunctional CD8[+] T cells by IL-2.**
**a**, Schematic of the experimental setup. One million Cor93 $T_N$ cells were transferred into HBV Tg mice. Indicated HBV Tg mice received IL-2c treatment 1 day after CD8[+] T cell transfer. Livers were collected and analysed 5 days after Cor93 $T_N$ cell transfer. Sera were collected before and 5 days after Cor93 $T_N$ cell transfer. **b**, Absolute numbers of IFNγ-producing Cor93 T cells in the livers of the indicated mice. $n = 3$ (control), 4 (IL-2c). **c**, ALT levels detected in the sera of the indicated mice. $n = 3$ (control), 4 (IL-2c). **d**, HBV DNA quantification (expressed as fold reduction over pre-treatment levels) in sera of the indicated mice before

and 5 days after Cor93 $T_N$ cell transfer. $n = 5$. **e**, HBV DNA quantification by Southern blot analysis in the liver of the indicated mice. Bands corresponding to the expected size of the integrated transgene, relaxed circular (RC), double-stranded linear (DS), and single-stranded (SS) HBV DNAs are indicated. $n = 5$ mice. **f**, Representative immunohistochemical micrographs of liver sections from the indicated mice showing HBV core antigen expression (brown). Scale bars, 100 μm. $n = 5$ mice. Data are mean ± s.e.m. and representative of at least two independent experiments. *$P < 0.05$, ***$P < 0.001$, two-tailed $t$-test. Mouse drawings were adapted from ref. [69].
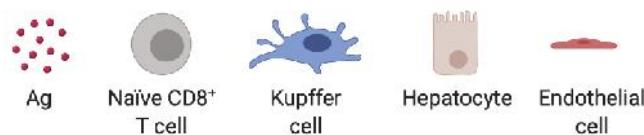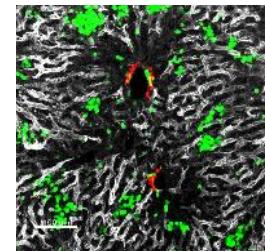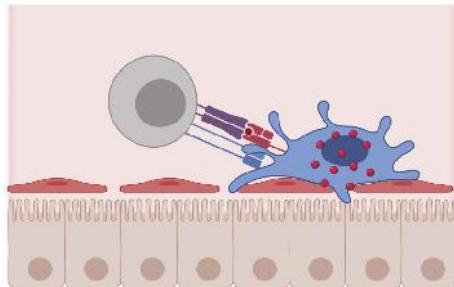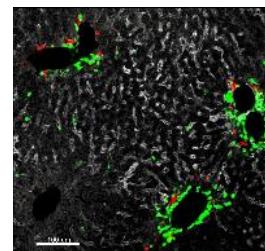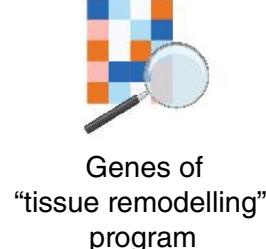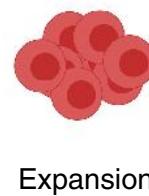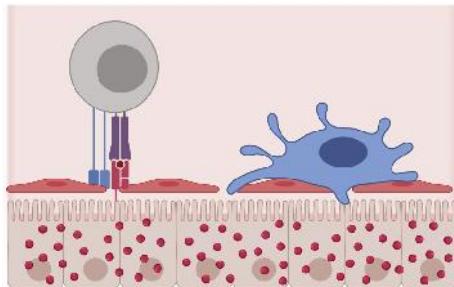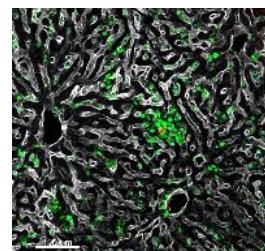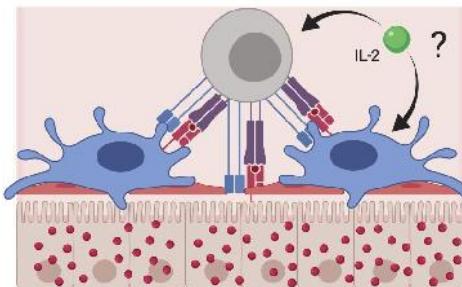
**Extended Data Fig. 10 | Summary of the main findings.** Top, priming by Kupffer cells—which are not natural targets of HBV—leads to differentiation into bona fide effector cells that form dense, extravascular clusters of rather immotile cells scattered throughout the liver. Middle, priming by hepatocytes—which are the natural targets of HBV—leads to local activation and proliferation but lack of differentiation into effector cells; these dysfunctional cells express a unique set of genes including some belonging to GO categories linked to tissue remodelling and they form loose, intravascular clusters of motile cells that coalesce around portal tracts. Bottom, CD8+ T cells primed by hepatocytes can be rescued by IL-2 treatment.

# nature research

Corresponding author(s): Matteo Iannacone

Last updated by author(s): Jul 30, 2019

# Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

## Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

| n/a | Confirmed | |
|---|---|---|
| ☐ | ☒ | The exact sample size ($n$) for each experimental group/condition, given as a discrete number and unit of measurement |
| ☐ | ☒ | A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly |
| ☐ | ☒ | The statistical test(s) used AND whether they are one- or two-sided<br>*Only common tests should be described solely by name; describe more complex techniques in the Methods section.* |
| ☐ | ☒ | A description of all covariates tested |
| ☐ | ☒ | A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons |
| ☐ | ☒ | A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals) |
| ☐ | ☒ | For null hypothesis testing, the test statistic (e.g. $F$, $t$, $r$) with confidence intervals, effect sizes, degrees of freedom and $P$ value noted<br>*Give P values as exact values whenever suitable.* |
| ☒ | ☐ | For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings |
| ☒ | ☐ | For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes |
| ☒ | ☐ | Estimates of effect sizes (e.g. Cohen's $d$, Pearson's $r$), indicating how they were calculated |

*Our web collection on statistics for biologists contains articles on many of the points above.*

## Software and code

Policy information about availability of computer code

| Data collection | BD FACSDiva 8, Imspector 6.4, ImageScope 12.3.2.8013, NineAlliance Mini9 |
|---|---|
| Data analysis | Prism 8, FlowJo 10, Imaris 9, Fiji , R 3.4.1, STAR aligner STAR_2.5.3a, BWA aligner 0.7.15-r1140, MACS2 2.1.1.20160309, BEDTools v2.24.0, SSAMtools 1.4, HOMER v4.10, bcl2fastq v2.20.0.422, Picard (MarkDuplicates) 1.104(1627), GSEA 3.0, Rsubrerad 1.24.2, edgeR 3.20.7, clusterProfiler 2.0.1, org.Mm.eg.db 3.7.0, CHIPpeakAnno 3.16.1 |

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

## Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

The RNA-seq and ATAC-seq data on sorted hepatic CD8+ T cells have been deposited in the ArrayExpress database under the accession code E-MTAB-7462 and EMTAB-7461, respectively. All data is available in the main text or the supplementary materials.

# Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

# Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

| | |
|---|---|
| Sample size | Sample sizes were chosen based on prior research conducted in our laboratories to provide sufficient numbers of mice in each group to provide informative results and perform statistical testing |
| Data exclusions | No data were excluded from analysis |
| Replication | Biological replicates were used to ensure reproducibility of this study. All presented data are representative of at least 2 independent experiments with similar results. All result described in the study could be reproduced. |
| Randomization | Mice were matched for age (4 or 8 weeks old), sex (Males) and (for the 1.3.32 animals) serum HBeAg levels and serum HBV DNA levels before randomization. |
| Blinding | Blinding was not performed as not relevant in this study, because subjective measurement was not involved. |

# Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

## Materials & experimental systems

| n/a | Involved in the study |
|---|---|
| ☐ | ☒ Antibodies |
| ☐ | ☒ Eukaryotic cell lines |
| ☒ | ☐ Palaeontology |
| ☐ | ☒ Animals and other organisms |
| ☐ | ☒ Human research participants |
| ☐ | ☒ Clinical data |

## Methods

| n/a | Involved in the study |
|---|---|
| ☒ | ☐ ChIP-seq |
| ☐ | ☒ Flow cytometry |
| ☒ | ☐ MRI-based neuroimaging |

## Antibodies

| | |
|---|---|
| Antibodies used | anti-CD3 (clone: 145-2C11, Cat#562286, BD Biosciences, 1:100), anti-CD11b (clone: M1/70, Cat#101239, BioLegend 1:100), anti-CD19 (clone: 1D3, Cat#562291 BD Biosciences, 1:100), anti-CD25 (clone: PC61, Cat#102015, BioLegend, 1:100), anti-CD31 (clone: 390, Cat#102427, BioLegend, 1:100), anti-CD45 (clone: 30-F11, Cat#564279 BD Biosciences, 1:100), anti-CD49b (clone: DX5, Cat#562453 BD Biosciences, 1:100), anti-CD64 (clone: X54-5/7.1, Cat#139311, BioLegend, 1:100), anti-F4/80 (clone: BM8, Cat#123117,BioLegend, 1:100), anti-I-A/I-E (clone: M5/114.15.2, Cat#107622, BioLegend, 1:100), anti-TIM4 (polyclonal, Cat#orb103599 Biorbyt, 1:100), anti-CD69 (clone: H1.2F3, Cat#104517, BioLegend, 1:100), anti-CD45.1 (clone: A20, Cat#110716,BioLegend, 1:100), anti-IFN-g (clone: XMG1.2, Cat#557735 BD Biosciences, 1:100), anti-CD4 (clone: RM4-5, Cat#553048 BD Biosciences, 1:100), anti-CD11c (clone: N418, Cat#117308, BioLegend, 1:100), anti-I-Ab (clone: AF6-120.1, Cat#116420, BioLegend, 1:100), anti-PD-1 (clone: J43, Cat#17-9985 eBioscience, 1:100), anti-NK1.1 (clone: PK136, Cat# 108706, BioLegend, 1:100), anti-NKp46 (clone: 29A1.4, Cat# 137623, BioLegend, 1:100), anti-Stat5 pY694 (clone: 47, Cat# 560117 BD Biosciences, 1:100), anti-Foxp3 (clone FJK-16s, Cat# 12-5773-80 eBioscience, 1:100), PB-conjugated anti-CD8a (clone 53-6.7, Cat#100725, BioLegend, 1:100), PE-conjugated anti-CD45.1 (clone: A20, Cat#110708, BioLegend, 1:100) |
| Validation | All antibodies were obtained from commercial vendors and we based specificity on descriptions and information provided in corresponding Data Sheets available and provided by the Manufacturers. Antibodies with high expression on intrahepatic leukocytes (e.g.CD45.1, CD69, CD3, CD4, CD8, CD45, IFN-g, CD25, etc.) have been previously used, validated, and published (see for instance Guidotti et al., Cell 2015). Representative flow panels were shown in Extended Data Fig.1f, 2h. |

# Eukaryotic cell lines

Policy information about <u>cell lines</u>

| Cell line source(s) | HEK-293T: A clone of human embryonic kidney 293 cells with stable transfection of the temperature-sensitive SV40 T-antigen was selected for its high-yield performance in production of LV, as described in Biffi et al., Science 2013. <br> CEM A3 cell line was obtained by American Type Culture Collection (ATCC). |
|---|---|
| Authentication | HEK-293T and CEM cell lines were not authenticated. |
| Mycoplasma contamination | HEK-293T and CEM cell lines are routinely tested for mycoplasma contamination before use and tested negative for mycoplasma contamination. |
| Commonly misidentified lines <br> (See <u>ICLAC</u> register) | Not listed in ICLAC |

# Animals and other organisms

Policy information about <u>studies involving animals</u>; <u>ARRIVE guidelines</u> recommended for reporting animal research

| Laboratory animals | C57BL/6, CD45.1 (inbred C57BL/6), Balb/c, Thy1.1 (CBy.PL(B6)-Thya/ScrJ), ?-actin-GFP [C57BL/6-Tg(CAG-EGFP)1Osb/J], ?-actin-DsRed [B6.Cg-Tg(CAG-DsRed*MST)1Nagy/J], Tap1-deficient (B6.129S2-Tap1tm1Arp/J), TCR-I [B6.Cg-Tg(TcraY1, TcrbY1)416Tev/J], CD11c-DTR [B6.FVB-1700016L2RikTg(Itgax-DTR/EGFP)57Lan/J] mice were purchased from Charles River or The Jackson Laboratory. MHC-II-/- mice were obtained through the Swiss Immunological Mutant Mouse Repository (Zurich, Switzerland). MUP-core transgenic mice (lineage MUP-core 50 [MC50], inbred C57BL/6, H-2b), that express the HBV core protein in 100% of the hepatocytes under the transcriptional control of the mouse major urinary protein (MUP) promoter, have been previously described{Guidotti:1994vj}. HBV replication-competent transgenic mice (lineage 1.3.32, inbred C57BL/6, H-2b), that express all of the HBV antigens and replicate HBV in the liver at high levels without any evidence of cytopathology, have been previously described{Guidotti:1995uf}. In indicated experiments, MUP-core and HBV replication-competent transgenic mice were used as C57BL/6 x Balb/c H-2bxd F1 hybrids. Cor93 TCR transgenic mice (lineage BC10.3, inbred CD45.1), in which > 98% of the splenic CD8+ T cells recognize a Kb-restricted epitope located between residues 93–100 in the HBV core protein (MGLKFRQL), have been previously described{Isogawa:2013ea}. Env28 TCR transgenic mice (lineage 6C2.36, inbred Thy1.1 Balb/c), in which ~83% of the splenic CD8+ T cells recognize a Ld-restricted epitope located between residues 28–39 of HBsAg (IPQSLDSWWTSL), have been previously described{Isogawa:2013ea}. For imaging experiments Cor93 and TCR-I transgenic mice were bred against both ?-actin-GFP and ?-actin-DsRed mice, while Env28 transgenic mice were bred against ?-actin-DsRed mice that were previously back-crossed more than 10 generations against Balb/c. Bone marrow (BM) chimeras were generated by irradiation of MUP-core or C57BL/6 mice with one dose of 900 rad and reconstitution with the indicated BM; mice were allowed to reconstitute for at least 4 weeks before use. In some experiments, to achieve full reconstitution of Kupffer cells from donor-derived BM, mice were injected with 200 ?l of clodronate-containing liposomes 42 and 40 days after BM injection. Mice were housed under specific pathogen-free conditions and used at 8-10 weeks of age. In all experiments, mice were matched for age (8 weeks old, 4 weeks old), sex (males animals) and for the 1.3.32 animals serum HBeAg levels before experimental manipulations. In selected experiments, 1.3.32 mice were matched for serum HBV DNA levels before experimental manipulations. |
|---|---|
| Wild animals | No wild animals were used in the study |
| Field-collected samples | No field collected samples were included in the study |
| Ethics oversight | All experimental animal procedures were approved by the Institutional Animal Committee of the San Raffaele Scientific Institute. |

Note that full information on the approval of the study protocol must also be provided in the manuscript.

# Human research participants

Policy information about <u>studies involving human research participants</u>

| Population characteristics | A total of 34 patients with chronic HBV infection (HBsAg+) were included. The patients were subdivided into the disease categories Immune Tolerant (IT), Immune Active (IA) based on their clinical history. The 13 IT patients had no history of hepatitis (normal ALT) and are all positive for HBeAg. The 16 IA patients (5 HBeAg+, 12 HBeAg-) have or had previously signs of hepatic inflammation (ALT > 40 IU/L), six of them are currently or were previously treated with nucleoside analogues. Supplementary Table 10 summarizes the available clinical and virological parameters. |
|---|---|
| Recruitment | Patients with chronic hepatitis B (HBsAg+ and HBV-DNA +) were recruited from the viral hepatitis clinic at The Royal London Hospital . Written informed consent was obtained from all subjects. The study was conducted in accordance with the Declaration of Helsinki and approved by the Barts and the London NHS Trust local ethics review board and the NRES Committee London–Research Ethics Committee (reference 10/H0715/39). |
| Ethics oversight | Written informed consent was obtained from all subjects. The study was conducted in accordance with the Declaration of Helsinki and approved by the Barts and the London NHS Trust local ethics review board and the NRES Committee London–Research Ethics Committee (reference 10/H0715/39) and by the Singapore National Healthcare Group ethical review board (DSRB 2008/00293). |

Note that full information on the approval of the study protocol must also be provided in the manuscript.

# Clinical data

All manuscripts should comply with the ICMJE guidelines for publication of clinical research and a completed CONSORT checklist must be included with all submissions.

| | |
|---|---|
| Clinical trial registration | The clinical study (not a clinical trial) was conducted in accordance with the Declaration of Helsinki and approved by the Barts and the London NHS Trust local ethics review board and the NRES Committee London–Research Ethics Committee (reference 10/H0715/39) and by the Singapore National Healthcare Group ethical review board (DSRB 2008/00293). |
| Study protocol | Patients were categorized into EASL 2017 standard phases using the clinical and virological criteria outlined in the EASL 2017 Clinical Practice Guidelines on the management of hepatitis B virus infection15: i) HBeAg+ chronic infection (eAg+CInf): normal ALT (< 40 IU/L), HBeAg positive and high HBV DNA; ii) HBeAg+ chronic hepatitis (eAg+CHep): elevated ALT, HBeAg positive; iii) HBeAg- chronic hepatitis (eAg-CHep): elevated ALT, anti-HBe positive; and iv) HBeAg- chronic infection (eAg-CInf): normal ALT, anti-HBe positive, low HBV DNA. Patients were followed for at least 1 year with virological and clinical parameters collected every 6 months. The ALT and virological parameters shown in Supplementary Table 10 are the ones present at the time of PBMC isolation. Patients were classified as immunotolerant (IT) if they had HBeAg+ chronic infection (eAg+CInf); alternatively, they were classified as immune active (IA) if they showed signs of immunological activity, that is eAg+CHep or eAg-CHep. |
| Data collection | Patients were followed for at least 1 year with virological and clinical parameters collected every 6 months. The ALT and virological parameters shown in Supplementary Table 10 are the ones present at the time of PBMC isolation. |
| Outcomes | The clinical study did not have outcomes (other than the results reported in this manuscript) as it is not a clinical trial. |

# Flow Cytometry

## Plots

Confirm that:

☐ The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).

☐ The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).

☒ All plots are contour plots with outliers or pseudocolor plots.

☒ A numerical value for number of cells or percentage (with statistics) is provided.

## Methodology

| | |
|---|---|
| Sample preparation | Sample preparation is described in the Materials & Methods section. |
| Instrument | BD FACS Canto or LSRII |
| Software | BD FACS DIVA for acquisition and FlowJo (Treestar) for analyses |
| Cell population abundance | Cells were always at least 98% pure |
| Gating strategy | Gating strategies is indicated in the Figure legends and the Materials & Methods section. |

☒ Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.

# ARTICLE

# Structural insights into the mechanism of human soluble guanylate cyclase

Yunlu Kang[1,4], Rui Liu[1,4], Jing-Xiang Wu[1] & Lei Chen[1,2,3]*

**Soluble guanylate cyclase (sGC) is the primary sensor of nitric oxide. It has a central role in nitric oxide signalling and has been implicated in many essential physiological processes and disease conditions. The binding of nitric oxide boosts the enzymatic activity of sGC. However, the mechanism by which nitric oxide activates the enzyme is unclear. Here we report the cryo-electron microscopy structures of the human sGCα1β1 heterodimer in different functional states. These structures revealed that the transducer module bridges the nitric oxide sensor module and the catalytic module. Binding of nitric oxide to the β1 haem–nitric oxide and oxygen binding (H-NOX) domain triggers the structural rearrangement of the sensor module and a conformational switch of the transducer module from bending to straightening. The resulting movement of the N termini of the catalytic domains drives structural changes within the catalytic module, which in turn boost the enzymatic activity of sGC.**

Nitric oxide (NO) is a gaseous signalling molecule that is involved in many important physiological processes, such as vasodilatation, neurotransmission, platelet aggregation, immunity, cell proliferation, and mitochondrial respiration[1,2]. The dysregulation of NO signalling has been linked to cardiovascular disease, sepsis, acute lung injury, and multiple organ failure[1,3,4]. NO signalling is initiated by the activation of NO synthase (NOS), which generates NO in response to physiological stimuli[4]. NO readily permeates target cell membranes, and after diffusing across the membrane, it binds and activates soluble guanylate cyclase (sGC), the primary NO acceptor[4]. sGC catalyses the cyclization reaction of guanosine triphosphate (GTP) to generate inorganic pyrophosphate and the secondary messenger cyclic guanosine monophosphate (cGMP)[5]. cGMP then acts on downstream effectors, including cGMP-regulated protein kinases, phosphodiesterases, and ion channels, to regulate physiological processes in the cell[5]. Genetic mutations of sGC in humans are associated with coronary artery disease[6], moyamoya disease, achalasia, and hypertension[7,8], and it is a validated drug target for the treatment of pulmonary hypertension and chronic heart failure[4]. The NO donor nitroglycerin has been widely used for centuries to alleviate angina pectoris, and the sGC stimulator riociguat has been approved for the treatment of pulmonary hypertension[4]. Drugs that activate or stimulate sGC also have therapeutic potential in fibrotic diseases, systemic sclerosis, chronic kidney diseases, neuroprotection, dementia, and sickle cell disease[9].

sGC is a heterodimeric protein complex composed of one α-subunit and one β-subunit. In humans, the α1 and β1 subunits are widely expressed in many tissues, while the expression of α2 and β2 subunits is tissue-specific[5,10]. The α- and β-subunits have some sequence homology and are similarly organized into modular domains, including an N-terminal H-NOX domain, a Per/Arnt/Sim (PAS) domain, a coiled-coil (CC) domain, and a C-terminal catalytic domain. The PAS and CC domains mediate protein–protein interactions, and the catalytic domain is responsible for enzymatic activity[5,11]. The H-NOX domain of the β-subunit contains a ferrous b-type haem prosthetic group that facilitates the high-affinity binding of NO[5,11]. Under pathological conditions or oxidative stress, the ferrous haem can be oxidized to ferric haem[12], and haem-oxidized sGC has low activity even in the presence of NO[13].

Several structures of isolated sGC domains have been solved by X-ray crystallography or NMR. These structures include the human β1 H-NOX domain (PDB ID: 5MNW), the *Manduca sexta* α-PAS domain[14], the human β1 CC domain[15], and the human α1β1 catalytic domain heterodimer[16,17]. Recent negative stain electron microscope studies[18] have revealed the general shape of the full-length mammalian sGC at a resolution of 25–40 Å, and hydrogen–deuterium exchange experiments mapped NO-induced structural changes onto the primary sequence of the full-length sGC[19]. Despite these pioneering structural efforts, the allosteric mechanism that underlies the activation of the distal catalytic domain in response to binding of NO to the β-subunit H-NOX domain remains unclear at the atomic level, mainly owing to the lack of high-resolution structural information on intact sGC in different functional states. Here, we have used cryo-electron microscopy (cryo-EM) to determine the structure of the human α1β1 sGC holoenzyme in both the inactive and NO-activated states at a resolution of 3.9 Å and 3.8 Å, respectively. We also obtained a 6.8 Å resolution cryo-EM map of the constitutively active β1(H105C) mutant. These structures uncover not only the detailed domain–domain interfaces, but also the activation mechanism of human sGC.

## Structure determination

We purified sGC composed of human α1 and β1 subunits to apparent homogeneity (Fig. 1a, Extended Data Fig. 1a, b), and the protein showed the characteristic ultraviolet–visual light (UV–vis) spectrum of sGC with ferrous haem bound (Extended Data Fig. 1c) and robust NO-activated GTP cyclase activity (Fig. 1b). By contrast, the β1(H105C) mutant, in which the haem group is unable to bind, showed constitutively high basal activity and was insensitive to NO activation[20] (Fig. 1b). We prepared a haem-unliganded state sample, a haem-oxidized state sample (Extended Data Fig. 1c) and an NO-activated state sample using the wild-type sGC protein (see Methods). These samples were subjected to single-particle cryo-EM analysis (Extended Data Figs. 1–4) and we obtained overall resolutions of 4 Å, 3.9 Å, and 3.8 Å, respectively. Map qualities were further improved by dividing the whole molecule into two bodies, the larger 'N lobe' and the smaller 'C lobe', and subsequent multibody refinements[21] (Supplementary

[1]State Key Laboratory of Membrane Biology, Institute of Molecular Medicine, Peking University, Beijing Key Laboratory of Cardiometabolic Molecular Medicine, Beijing, China. [2]Peking-Tsinghua Center for Life Sciences, Peking University, Beijing, China. [3]Academy for Advanced Interdisciplinary Studies, Peking University, Beijing, China. [4]These authors contributed equally: Yunlu Kang, Rui Liu. *e-mail: chenlei2016@pku.edu.cn
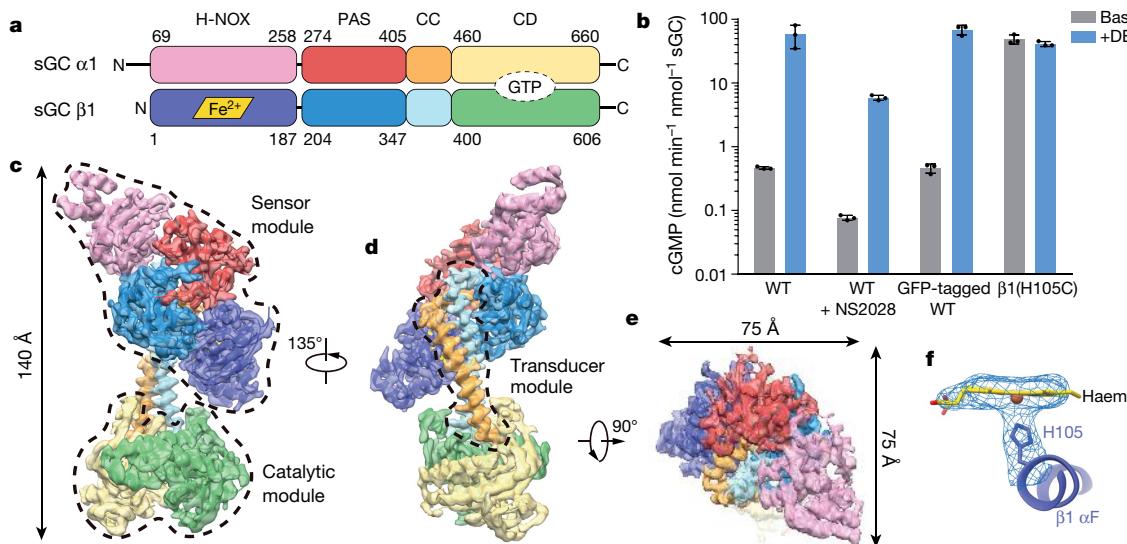
**Fig. 1 | Structure of sGC in the inactive state. a**, Domain organization of the human α1β1 sGC heterodimer. CD, catalytic domain. The haem cofactor and GTP substrate binding site are shown as a parallelogram and an oval, respectively. **b**, End-point activity assay of the various sGC protein samples with or without DEA. WT, wild-type. NS2028 oxidizes the Fe(II) in sGC to Fe(III). Mean ± s.d., $n = 3$ biologically independent samples. **c**, Side view of the cryo-EM map of sGC in the inactive (haem-oxidized)

state. The colours of each individual domain are as in **a**. The atomic model is shown as a cartoon inside the transparent electron density surface. The approximate boundaries of the sensor module and the catalytic module are shown in dashed lines. **d**, A 135° rotated view compared to **c**. The approximate boundary of the transducer module is shown in dashed lines. **e**, A 90° rotated top view compared to **d**. **f**, The cryo-EM density map of the β1 H105–haem region in the inactive (haem-oxidized) sGC.

Videos 1, 2). The qualities of composite maps obtained from multibody refinement were sufficient to trace the main chain of most residues with the aid of available high-resolution homologous structures (Extended Data Figs. 5, 6, Extended Data Table 1). We also obtained a 6.8 Å map for the β1(H105C) mutant (Extended Data Fig. 4g, h). At this resolution, the overall shape and domain organization of the β1(H105C) mutant were found to be similar to that of the NO-activated state, with a real space correlation of 0.96. However, the haem density is missing in the β1 H-NOX domain of the β1(H105C) mutant, as expected (Extended Data Fig. 4i). The atomic models of sGC in different states allowed us to characterize the domain–domain interfaces in detail (Extended Data Figs. 7, 8).

## Structure of sGC in the inactive state

Both the haem-unliganded and the haem-oxidized sGC were in a 'bent' conformation[18] (Fig. 1c–e, Supplementary Video 3). In our cryo-EM reconstructions, we found that the overall structure of sGC in the haem-unliganded state (4 Å) is essentially the same as that in the haem-oxidized state (3.9 Å), with a root mean square deviation (r.m.s.d.) of only 0.28 Å (Extended Data Fig. 9a), in accordance with the functional data, which showed that the haem-unliganded and haem-oxidized states have low activity[13] (Fig. 1b). Therefore, both of the structures were considered as the inactive state, and the 3.9 Å haem-oxidized state is used in further discussion of the inactive state. The structure of the inactive sGC occupies a 3D space of 140 Å × 75 Å × 75 Å (Fig. 1c–e). The large N lobe is composed of α1 H-NOX, α1 PAS, β1 PAS, and β1 H-NOX domains. These domains are arranged in a pseudo-two-fold symmetric manner, with the scaffolding PAS domains at the centre and the H-NOX domains at the periphery (Fig. 1c). The H-NOX domains and PAS domains are essential for NO sensing and form the N-terminal sensor module of sGC (Fig. 1c). The CC domains of both subunits form the transducer module that bridges the N-terminal sensor module and the C-terminal catalytic module (Fig. 1d).

A haem molecule binds inside the β1 H-NOX domain, and its five-coordinated Fe ion is tightly bound to H105 of αF, as evidenced by the strong connecting density between them (Fig. 1f). By contrast, the α1 H-NOX domain does not bind haem owing to a sterical clash (Extended Data Fig. 9b). The structure of each PAS domain resembles that of the *M. sexta* sGC α-subunit[14] (PDB ID: 4GJ4; Extended Data

Fig. 9c, d). Extensive domain–domain interactions stabilize the structure of sGC in the inactive state (Extended Data Fig. 7, Supplementary Notes 1–3). The β1 H-NOX domain, especially the αE and αF helices, interacts with both the neighbouring PAS heterodimers and the transducer module (Extended Data Figs. 2g, 7a–e, Supplementary Note 1). These interactions further stabilize the transducer module in the bent conformation, in which both the α1 and β1 CC domains are broken into two short helices (αM and αN) connected by a near 90° turn (Extended Data Fig. 7f). The two αN helices pack in a 'leucine zippers' manner and interact extensively with the catalytic module (Extended Data Fig. 7f–h, Supplementary Note 2). In the catalytic module, the two subunits are organized in a pseudo-symmetric manner as well, but the angle between domains is different from that of the isolated catalytic module[17] (Extended Data Fig. 9e, f, Supplementary Note 3). Compared with the adenylate cyclase in the active state[22] (PDB ID: 1CJU), the structure of the catalytic module shows steric clashes between the substrate and the protein residues (Extended Data Fig. 9g). This suggests that the structure of sGC in the inactive state is incompatible with substrate binding, consistent with previous studies that showed that inactive sGC has a high Michaelis constant ($K_m$)[23]. The domain–domain interactions observed in the inactive state were further validated by cysteine cross-linking under oxidative conditions (Extended Data Fig. 7i–l, Supplementary Note 4).

## Structure of sGC in NO–activated state

The NO-activated sGC has a dumbbell-shape extended structure[18], in which the sensor module moves away from the catalytic module (Fig. 2, Supplementary Video 4). This is markedly different from the bent conformation of the inactive state. In addition, the overall structure of the constitutively active β1(H105C) mutant, in the absence of NO donors, shows a similar extended conformation (Extended Data Fig. 4g). This structural agreement suggests that this large overall conformational change is associated with enhanced enzymatic activity and the full activation of sGC, but probably does not result from the *S*-nitrosylation of sGC by NO, which is a covalent modification of cysteine residues that can lead to desensitization of sGC under certain conditions[24–26] (Extended Data Fig. 4j, Supplementary Note 5).

Despite the large overall conformational change, the general domain arrangement within each module in the NO-activated state is
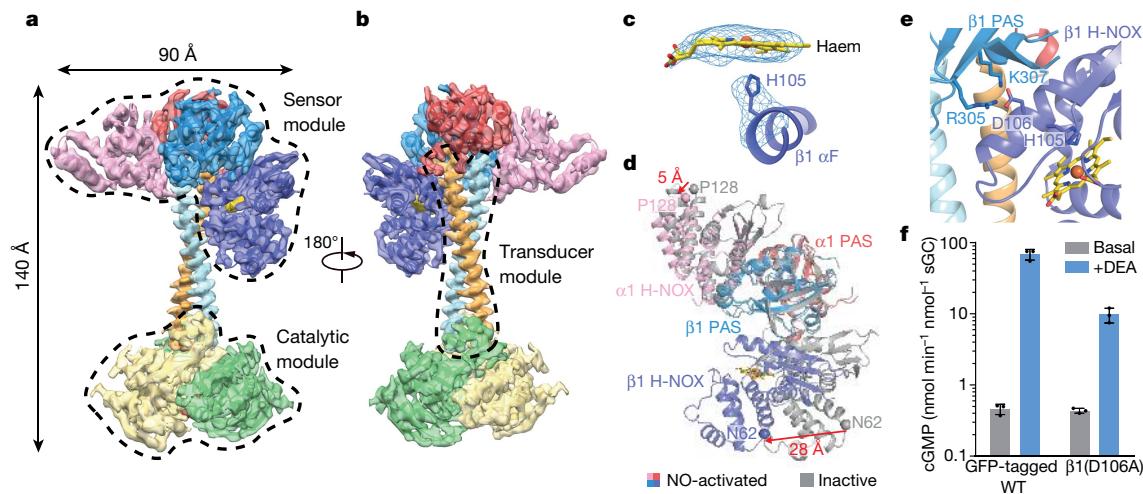
**Fig. 2 | Structure of sGC in the NO-activated state. a**, Side view of the cryo-EM map of sGC in the NO-activated state. The colours of each individual domain are the same as in Fig. 1a. **b**, A 180° rotated view compared to **a**. **c**, The cryo-EM density map of the β1 H105–haem region in the NO-activated sGC. **d**, Structural rearrangement of the sensor module. The structure of sGC in the NO-activated state (coloured as in Fig. 1a) is superimposed onto the structure of sGC in the inactive state (grey) by aligning the PAS heterodimers. The arrows indicate the positional changes of the Cα atoms of α1 P128 and β1 N62 induced by NO binding. **e**, The interface between β1 H-NOX and the adjacent β1 PAS domain of sGC in the NO-activated state. Polar interactions involving D106 are highlighted by showing their side chains as sticks. **f**, End-point activity assay of the β1 subunit mutant. Mean ± s.d., $n = 3$ biologically independent samples.

maintained (Figs. 1, 2). In the electron density map of the NO-activated state, the H105–Fe bond of β1 H-NOX is cleaved, as evidenced by the clear separation between each density (Fig. 2c). This suggests that the current conformation is likely to correspond to an NO-bound state, because excess NO donor DEA NONOate was added to the sample and NO binds sGC with picomolar-range high affinity[2]. However, we could not explicitly model the NO molecules or the haem deformation owing to the limited resolution. The binding of NO induces a conformational change in β1 H-NOX in which the C-terminal subdomain rotates relative to the N-terminal subdomain (Extended Data Fig. 9h). When αF (residues 96–107) of the β1 subunit was used as the reference to superimpose the structure of the NO-bound β1 H-NOX domain onto the structure of the inactive state, the Cα atom of N62 in the N-terminal subdomain was displaced by 4.6 Å (Extended Data Fig. 8a) and, more importantly, the NO-bound β1 H-NOX domain sterically clashed with the adjacent domains of the inactive state (Extended Data Fig. 8a). This indicates that the inactive state structure is incompatible with the NO-bound β1 H-NOX domain and, therefore, a structural rearrangement is required to accommodate the conformational change of the β1 H-NOX domain upon NO binding. Indeed, we observed structural changes within the sensor module in which α1 H-NOX underwent a small downward movement while β1 H-NOX underwent a large rotational and translational movement (Fig. 2d).

These conformational changes of the sensor module in the NO-activated state result in completely new interfaces between the NO-bound β1 H-NOX domain and its adjacent domains (Extended Data Figs. 4e, 8b, c, Supplementary Notes 6, 7). Many residues contribute to this new interface; among them, D106 on αF of the β1 H-NOX domain forms important polar interactions with other residues (Fig. 2e). We found that sGC with the β1(D106A) mutation had normal haem incorporation but impaired activation by NO (Fig. 2f, Extended Data Fig. 8d), suggesting that this interface is essential for sGC activity in the NO-activated state.

The markedly altered interactions between the CC domain and the sensor module lead to the conformational change of the transducer module (Fig. 3). Strikingly, the linkers between αM and αN fold into α-helical structures, which fuse αM with αN into the 71 Å-long αMN helices (Extended Data Fig. 8e). Specifically, R420–K426 of the α1 CC domain and L355–Q358 of the β1 CC domain fold into α-helical structures (Fig. 3a, Extended Data Figs. 5, 6). As a result, the transducer module switches from a highly bent conformation in the inactive state

into a long, continuous coiled-coil structure in the NO-activated state (Fig. 3a). The folding of the αM–αN loops results in a decrease in the exchangeability of the main chain hydrogens as they form hydrogen bonds in α-helices. This is consistent with previous hydrogen–deuterium exchange mass spectrum results that showed that the αM–αN loops had a much slower exchange rate upon NO activation[19]. To determine the functional importance of this bending–straightening conformational change, we mutated residues in the αM–αN linker to either prolines or alanines. Prolines generate kinks in helical structures because they cannot form hydrogen bonds on the main chain. Therefore, proline mutations should destabilize the helical structures of αMNs in the NO-activated state, and these proline mutants may favour the inactive conformation. Indeed, proline mutations of D423 in the α1 CC domain or G356 in the β1 CC domain rendered sGC unresponsive to NO activation, although these mutants could incorporate haem normally (Fig. 3a, b, Extended Data Fig. 8d). By contrast, mutations of the same set of residues into alanines had no such effect (Fig. 3a, b), indicating that the continuous helical structures of the αMNs are essential for activation of sGC by NO.

In the NO-activated conformation, the interface between the α1 and β1 CC domains is markedly different from that observed in the inactive state (Fig. 3a, Extended Data Fig. 7f, Supplementary Note 8). Besides the overall bending–straightening movements of each CC domain, the αN helix of the α1 subunit rotates approximately 70° around the αN helix of the β1 subunit (Fig. 3c). The separation of the C termini of the transducer modules also decreases. The distance between the Cα atoms of P459 of the α1 subunit and P399 of the β1 subunit shrinks from 26 Å to 20 Å (Fig. 3d). This drives the structural reorganization of the connecting catalytic module, in which the catalytic domain of the α1 subunit rotates 17° relative to the β1 catalytic domain (Fig. 3e, Extended Data Figs. 8f, 9i, j). These movements increase the volume of the central pocket from 1,375 Å³ to 1,549 Å³ and reorganize the catalytic centre (Extended Data Fig. 9k). This not only permits the binding of the substrate GTP and the cofactor $Mg^{2+}$ ions but also alters the local chemical environment of the pocket to make it possible for small stimulators to plug in and activate the enzyme (Supplementary Note 9). In the map of the NO-activated state, we observe a strong density corresponding to the substrate analogue GMPCPP (Fig. 3f), which was added during cryo-EM sample preparation. By comparing the current structure with the active adenylate cyclase structure[22] (PDB ID: 1CJU) (Extended Data Fig. 9l), we found that the residues responsible for substrate binding
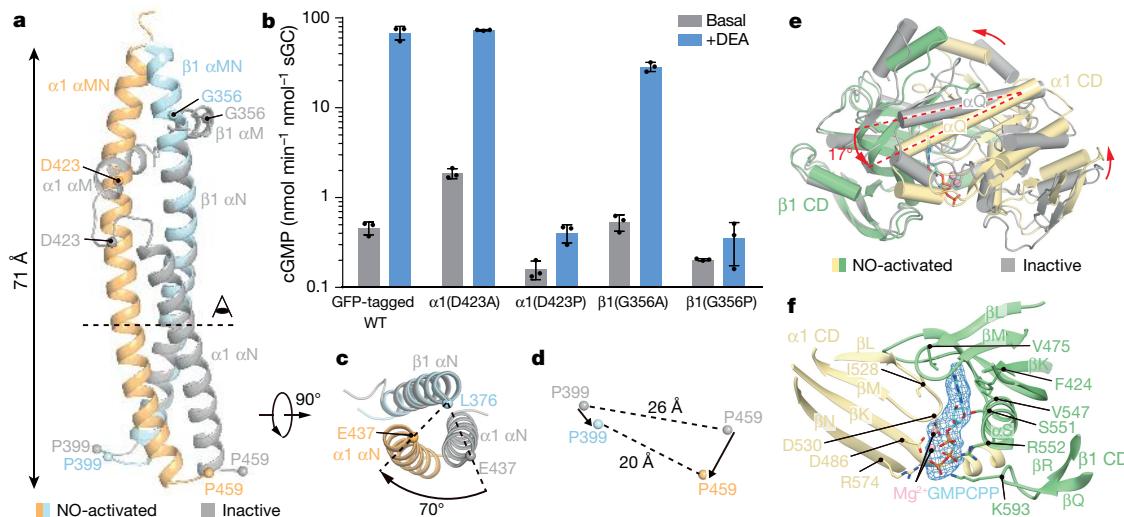
**Fig. 3 | The structure of the sGC transducer and catalytic module in the NO-activated state. a**, Structural comparison of the transducer module in the inactive state (grey) and the NO-activated state (coloured) by aligning the αN helices of their β1 subunits. The positions of Cα atoms of several key residues in the αM–αN linker are shown as spheres. **b**, End-point activity assay of the proline and alanine mutants in the αM–αN loop. Mean ± s.d., n = 3 biologically independent samples. **c**, A 90° rotated top view of **a**, beginning at the plane indicated with a dashed line and the point of view in **a**. The angle between the Cα atoms of α1 E437 was measured using the Cα atom of β1 L376 as the vertex. **d**, Positional displacement

of the Cα atoms of β1 P399 and α1 P459. The Cα atoms in the inactive and NO-activated state are shown as grey spheres and coloured spheres, respectively. The arrows denote the direction of change during activation. **e**, Side view of the structural comparison of the catalytic module between the inactive state (grey) and the NO-activated state (coloured). The GMPCPP molecule is shown as sticks. The angle between the αQ helices is shown. **f**, The structure of the sGC catalytic core in the NO-activated state. GMPCPP is shown as cyan sticks and magnesium ions are shown as pink spheres. The densities of GMPCPP and magnesium are shown as blue mesh.

and catalysis are in similar positions, indicating that the current sGC structure represents a catalytically competent conformation.

## Structural mechanism of sGC activation

By analysing the structures of individual sGC domains in both inactive and NO-activated states, we found that the structures of the scaffolding PAS dimer remain relatively unchanged among the different states, with a r.m.s.d. of 0.91 Å (Extended Data Fig. 9m). Therefore, we used the structures of the PAS dimer as a reference point to align and compare the two full-length structures (Fig. 4a). During activation, the α1 H-NOX domain makes a small concomitant downward movement, while the interfaces between α1 H-NOX and its adjacent domains are largely maintained (Fig. 2d). This suggests that the α1 H-NOX domain may play a role that is mainly structural instead of being involved in NO signal transduction (Supplementary Video 5). This is in agreement with the finding from an activity assay that the H-NOX domain of the α1 subunit is dispensable for NO activation[27] (Fig. 4b). By contrast,

the local conformational change of the β1 H-NOX domain upon NO binding drives the structural rearrangement of the sensor module (Fig. 2d), which, along with previous functional data, suggests that the H-NOX domain of the β1 subunit has an essential role in NO sensing[5]. Indeed, complete removal of the β1 H-NOX domain rendered the sGC enzyme trapped in a relatively low activity state and unresponsive to NO activation (Fig. 4b). This suggests that the β1 H-NOX domain in the NO-bound state is necessary to stabilize the sGC enzyme in an active conformation. Further supporting this conclusion, disruption of the interactions between β1 H-NOX and adjacent domains by mutation also diminished activation by NO (Fig. 2f). The structural changes in the sensor module upon binding of NO trigger the bending–straightening conformational switch of the transducer module. As a result, the distal catalytic module rotates 86° in a swing-like manner and the centre of mass of the catalytic module is spatially displaced by 101 Å (Fig. 4a). Inhibition of the straightening by proline mutations abolishes activation by NO (Fig. 3b), which suggests that the conformational
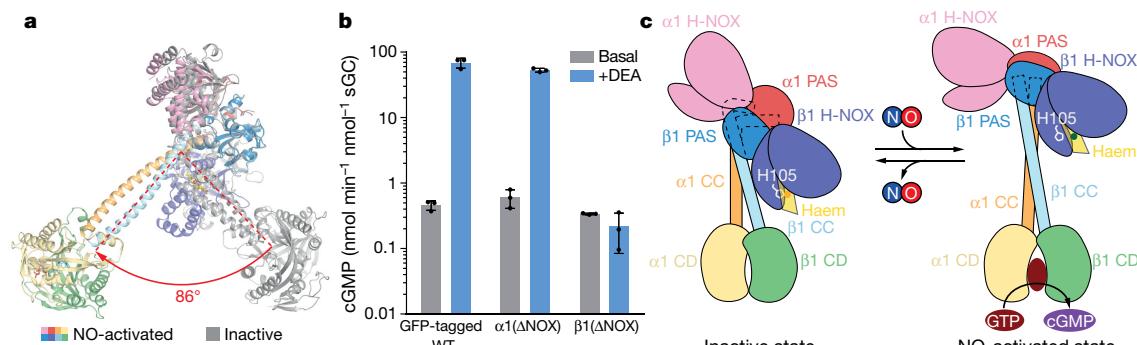


**Fig. 4 | Overall structural rearrangement of sGC during NO activation. a**, The structure of sGC in the NO-activated state (coloured as in Fig. 1a) is overlaid onto the structure of sGC in the inactive state (grey) by superimposing the PAS domain dimer. The angle between the αN helices of their β1 subunits is shown below. **b**, End-point activity assay

of the H-NOX domain deletion mutants. Mean ± s.d., n = 3 biologically independent samples. **c**, Cartoon model of the conformational changes during sGC activation. The colours of each individual domain are as in Fig. 1a.

change in the transducer module is essential for activation of the catalytic module. Furthermore, we found that the catalytic module changes from a conformation that cannot bind substrate to a catalytically competent conformation (Fig. 3e, f, Extended Data Figs. 8f, 9g, k), which explained how the binding of NO decreases the $K_{m (GTP)}$ and increases the catalytic constant ($k_{cat}$) of sGC[23] (Fig. 4c). It has been previously proposed that the activation of sGC by NO involves two steps[2], and our structure observations are compatible with this two-step hypothetic model (Supplementary Note 10). Notably, information flow in the reverse direction, from the catalytic module to the sensor module, has been suggested by published functional data[28–30]. Therefore, the transducer module acts as an allosteric structural coupler between the sensor module and the catalytic module to allow the bi-directional flow of information within the sGC molecule (Supplementary Note 11).

## Online content

1. Hollenberg, S. M. & Cinel, I. Bench-to-bedside review: nitric oxide in critical illness—update 2008. *Crit. Care* **13**, 218 (2009).
2. Horst, B. G. & Marletta, M. A. Physiological activation and deactivation of soluble guanylate cyclase. *Nitric Oxide* **77**, 65–74 (2018).
3. Luiking, Y. C., Engelen, M. P. & Deutz, N. E. Regulation of nitric oxide production in health and disease. *Curr. Opin. Clin. Nutr. Metab. Care* **13**, 97–104 (2010).
4. Farah, C., Michel, L. Y. M. & Balligand, J. L. Nitric oxide signalling in cardiovascular health and disease. *Nat. Rev. Cardiol.* **15**, 292–316 (2018).
5. Derbyshire, E. R. & Marletta, M. A. Structure and regulation of soluble guanylate cyclase. *Annu. Rev. Biochem.* **81**, 533–559 (2012).
6. Deloukas, P. et al. Large-scale association analysis identifies new risk loci for coronary artery disease. *Nat. Genet.* **45**, 25–33 (2013).
7. Hervé, D. et al. Loss of α1β1 soluble guanylate cyclase, the major nitric oxide receptor, leads to moyamoya and achalasia. *Am. J. Hum. Genet.* **94**, 385–394 (2014).
8. Wallace, S. et al. Disrupted nitric oxide signaling due to GUCY1A3 mutations increases risk for moyamoya disease, achalasia and hypertension. *Clin. Genet.* **90**, 351–360 (2016).
9. Sandner, P. From molecules to patients: exploring the therapeutic role of soluble guanylate cyclase stimulators. *Biol. Chem.* **399**, 679–690 (2018).
10. Koesling, D., Mergia, E. & Russwurm, M. Physiological functions of NO-sensitive guanylyl cyclase isoforms. *Curr. Med. Chem.* **23**, 2653–2665 (2016).
11. Montfort, W. R., Wales, J. A. & Weichsel, A. Structure and activation of soluble guanylyl cyclase, the nitric oxide sensor. *Antioxid. Redox Signal.* **26**, 107–121 (2017).
12. Dasgupta, A., Bowman, L., D'Arsigny, C. L. & Archer, S. L. Soluble guanylate cyclase: a new therapeutic target for pulmonary arterial hypertension and chronic thromboembolic pulmonary hypertension. *Clin. Pharmacol. Ther.* **97**, 88–102 (2015).
13. Zhao, Y. et al. Inhibition of soluble guanylate cyclase by ODQ. *Biochemistry* **39**, 10848–10854 (2000).
14. Purohit, R., Weichsel, A. & Montfort, W. R. Crystal structure of the Alpha subunit PAS domain from soluble guanylyl cyclase. *Protein Sci.* **22**, 1439–1444 (2013).
15. Ma, X., Beuve, A. & van den Akker, F. Crystal structure of the signaling helix coiled-coil domain of the β1 subunit of the soluble guanylyl cyclase. *BMC Struct. Biol.* **10**, 2 (2010).
16. Allerston, C. K., von Delft, F. & Gileadi, O. Crystal structures of the catalytic domain of human soluble guanylate cyclase. *PLoS ONE* **8**, e57644 (2013).
17. Seeger, F. et al. Interfacial residues promote an optimal alignment of the catalytic center in human soluble guanylate cyclase: heterodimerization is required but not sufficient for activity. *Biochemistry* **53**, 2153–2165 (2014).
18. Campbell, M. G., Underbakke, E. S., Potter, C. S., Carragher, B. & Marletta, M. A. Single-particle EM reveals the higher-order domain architecture of soluble guanylate cyclase. *Proc. Natl Acad. Sci. USA* **111**, 2960–2965 (2014).
19. Underbakke, E. S. et al. Nitric oxide-induced conformational changes in soluble guanylate cyclase. *Structure* **22**, 602–611 (2014).
20. Martin, E., Sharina, I., Kots, A. & Murad, F. A constitutively activated mutant of human soluble guanylyl cyclase (sGC): implication for the mechanism of sGC activation. *Proc. Natl Acad. Sci. USA* **100**, 9208–9213 (2003).
21. Nakane, T., Kimanius, D., Lindahl, E. & Scheres, S. H. Characterisation of molecular motions in cryo-EM single-particle data by multi-body refinement in RELION. *eLife* **7**, e36861 (2018).
22. Tesmer, J. J. et al. Two-metal-ion catalysis in adenylyl cyclase. *Science* **285**, 756–760 (1999).
23. Denninger, J. W. et al. Interaction of soluble guanylate cyclase with YC-1: kinetic and resonance Raman studies. *Biochemistry* **39**, 4191–4198 (2000).
24. Sayed, N., Baskaran, P., Ma, X., van den Akker, F. & Beuve, A. Desensitization of soluble guanylyl cyclase, the NO receptor, by S-nitrosylation. *Proc. Natl Acad. Sci. USA* **104**, 12312–12317 (2007).
25. Mayer, B. et al. Inactivation of soluble guanylate cyclase by stoichiometric S-nitrosation. *Mol. Pharmacol.* **75**, 886–891 (2009).
26. Fernhoff, N. B., Derbyshire, E. R., Underbakke, E. S. & Marletta, M. A. Heme-assisted S-nitrosation desensitizes ferric soluble guanylate cyclase to nitric oxide. *J. Biol. Chem.* **287**, 43053–43062 (2012).
27. Koglin, M. & Behrends, S. A functional domain of the α1 subunit of soluble guanylyl cyclase is necessary for activation of the enzyme by nitric oxide and YC-1 but is not involved in heme binding. *J. Biol. Chem.* **278**, 12590–12597 (2003).
28. Ruiz-Stewart, I. et al. Guanylyl cyclase is an ATP sensor coupling nitric oxide signaling to cell metabolism. *Proc. Natl Acad. Sci. USA* **101**, 37–42 (2004).
29. Cary, S. P., Winger, J. A. & Marletta, M. A. Tonic and acute nitric oxide signaling through soluble guanylate cyclase is mediated by nonheme nitric oxide, ATP, and GTP. *Proc. Natl Acad. Sci. USA* **102**, 13064–13069 (2005).
30. Tsai, A. L., Berka, V., Sharina, I. & Martin, E. Dynamic ligand exchange in soluble guanylyl cyclase (sGC): implications for sGC regulation and desensitization. *J. Biol. Chem.* **286**, 43182–43192 (2011).

## METHODS

**Cell culture.** HEK293F (Thermo Fisher Scientific) suspension cells were cultured in Freestyle 293 medium (Thermo Fisher Scientific) or SMM 293-TI medium (Sino Biological) supplemented with 1% FBS at 37 °C with 6% $CO_2$ and 70% humidity. It is reported that HEK293F is a female cell line. Sf9 insect cells (Thermo Fisher Scientific) were cultured in SIM SF (Sino Biological) at 27 °C. The cell lines were routinely checked to be negative for mycoplasma contamination but have not been authenticated.

**Protein expression and purification.** cDNA of *Drosophila melanogaster*[31], mouse, and human sGC were cloned into a modified BacMam expression vector[32,33] and transfected into HEK293F cells for screening by fluorescence-detection size-exclusion chromatography (FSEC)[34] on a Superose 6 increase 5/150 GL. The combination of C-terminal GFP-tagged human α1 and non-tagged β1 subunits yielded a stable heterodimer. sGC protein composed of an α1 and a β1 subunit is the most predominant isoform, and it has been widely used as a model protein to elucidate the biochemical, biophysical, and structural properties of mammalian sGC[5]. The coding sequences of human α1 and β1 subunits were transformed into the pFastbac dual vector and expression was driven by p10 or polyhedrin promoters. The corresponding baculovirus was generated using the Bac-to-Bac system.

Sf9 insect cells at a density of $4 \times 10^6$/ml in SIM SF medium were infected with the baculovirus and cultured at 27 °C in a shaker for 72 h before harvesting and storage at −80 °C. Cells corresponding to 500 ml culture were thawed and resuspended with 20 ml lysis buffer (50 mM Tris pH 8.0 at 4 °C, 150 mM NaCl) containing 1 µg/ml aprotinin, 1 µg/ml pepstatin, 1 µg/ml leupeptin, 1 mM phenylmethanesulfonyl fluoride (PMSF), 2 mM dithiothreitol (DTT), and 1 mM ethylenediaminetetraacetic acid (EDTA). Cells were broken by sonication in 5 s intervals followed by a 5 s pause at 50% output for 20 min. Unbroken cells, cell debris, and membranes were removed by ultracentrifugation at 40,000 r.p.m. for 1 h at 4 °C using a Ti70 rotor (Beckman). An excess amount of purified glutathione *S*-transferase-tagged GFP-nanobody[35] was added to the supernatant and incubated at 4 °C for 10 min with rotation. Samples were then loaded onto 4-ml Glutathione Sepharose 4B columns (GE Healthcare) and washed with TBS buffer (20 mM Tris, pH 8.0, 150 mM NaCl) containing 1 mM DTT at 4 °C. Protein was eluted with elution buffer (50 mM Tris, pH 8.5, 10 mM reduced glutathione, 1 mM DTT) at 4 °C. The eluate was diluted with buffer A (20 mM Tris, pH 8.0 at 4 °C) to a conductivity lower than 5 mS/cm and loaded onto a 1-ml HiTrap Q HP (GE Healthcare). The protein was eluted with buffer B (20 mM Tris, pH 8.0, 500 mM NaCl) at 4 °C in a linear gradient using the AKTA pure system (GE Healthcare). The peak fractions containing sGC were pooled and incubated with prescission protease overnight to cleave the tag from the protein. The digested protein was further purified by Superdex 200 increase (GE Healthcare) running in buffer containing 20 mM HEPES (pH 7.4), 50 mM NaCl and 2 mM tris (2-carboxyethyl) phosphine (TCEP). The peak fractions containing the sGC protein were pooled and concentrated. UV-vis spectrum was measured using a spectrometer (Pultton) in the cuvette mode.

**Activity assay.** The protein used for cryo-EM sample preparation was diluted with 20 mM triethanolamine (TEA, pH 7.6), 300 mM NaCl, 1 mM DTT and subjected to activity assay as described below. For the haem-oxidized sample, the protein was diluted with 20 mM TEA (pH 7.6), 300 mM NaCl and preincubated with 20 µM NS2028 at 25 °C for 30 min and then added DTT to the final concentration of 1 mM for activity assay. To generate the sGC mutant protein for activity assay, the coding sequences of the α1 subunit with a C-terminal GFP-strep tag and the β1 subunit were cloned into pFastBac1 expression vectors, respectively. To generate α1(ΔNOX) and β1(ΔNOX) constructs, the N-terminal 273 amino acids of the α1 subunit and 200 amino acids of the β1 subunit were removed, respectively. Constructs carrying the desired point mutations were generated by Quick Change and corresponding baculoviruses were generated using the Bac-to-Bac system. Sf9 insect cells at a density of $4 \times 10^6$/ml in SIM SF medium were infected with baculovirus and cultured in a shaker at 27 °C for an additional 72 h before harvest. Cells were resuspended with buffer containing 50 mM Tris (pH 8.0 at 4 °C), 150 mM NaCl, 1 µg/ml aprotinin, 1 µg/ml pepstatin, 1 µg/ml leupeptin, 1 mM PMSF and 1 mM DTT, and broken by passing through a syringe needle with a 0.45-µm inner diameter six times. Cell debris and membrane were removed by ultracentrifugation at 40,000 r.p.m. for 30 min at 4 °C using a TLA55 rotor (Beckman). The supernatants were loaded onto Streptactin Beads 4 FF (Smart-Lifesciences) and washed with buffer containing 20 mM Tris (pH 8.0 at 4 °C), 150 mM NaCl and 1 mM DTT. The protein was eluted with buffer containing 50 mM Tris (pH 8.5 at 4 °C), 50 mM NaCl, 10 mM D-desthiobiotin and 1 mM DTT. The eluates were diluted with an equal volume of 20 mM TEA (pH 7.6), loaded onto Q Beads 6 FF (Smart-Lifesciences) by gravity and washed with buffer containing 20 mM TEA (pH 7.6), 150 mM NaCl and 1 mM DTT. The protein was eluted by 20 mM TEA (pH 7.6), 300 mM NaCl and 1 mM DTT. The protein concentrations of various GFP-tagged sGC mutants were determined by comparing their GFP fluorescence signal to that of a purified GFP-tagged sGC standard on FSEC[34]. The activity assay

mixture contained 10 nM sGC, 60 mM TEA (pH 7.6), 150 mM NaCl, 0.5 mM DTT, 5 mM $MgCl_2$, 200 µM GTP with or without 200 µM DEA NONOate (Cayman Chemical) in a final volume of 20 µl. The assay mixture was incubated at 25 °C for 10 min and stopped by adding 80 µl 125 mM $Zn(OAc)_2$ and 100 µl 125 mM $Na_2CO_3$. The GTP-$ZnCO_3$ precipitation was removed by centrifugation at 17,000*g* for 5 min and the supernatant was used for cGMP quantification with the Cyclic GMP ELISA Kit (Cayman Chemical) according to the instructions. Each assay was independently repeated at least three times. For the measurement of UV-vis spectra of sGC mutants, proteins eluted from Streptactin Beads 4FF (Smart-Lifesciences) were digested with prescission protease overnight and further purified by Superdex 200 increase column (GE Healthcare) with buffer containing 20 mM HEPES (pH 7.4), 50 mM NaCl and 1 mM TCEP. The peak fractions were pooled and concentrated. UV-vis spectrums of sGC mutants with or without 400 µM DEA NONOate were measured using a spectrometer (Pultton).

**EM sample preparation.** We prepared a haem-unliganded sGC sample, in which no exogenous ligand was supplemented, and then supplemented different small molecules to stabilize the purified sGC protein in functionally distinct states. The compound NS2028 has been reported to efficiently oxidize the Fe(II) in sGC to Fe(III)[36]. Indeed, we found that incubating sGC with NS2028 almost completely shifted the Soret peak from 431 to 392 nm (Extended Data Fig. 1c). Therefore, we incubated purified sGC with NS2028, $Mg^{2+}$ ions, and substrate GTPγS[37] to obtain the haem-oxidized state. To achieve the NO-activated state, we supplemented the purified protein with excess NO donor DEA NONOate[38], $Mg^{2+}$ ions, and noncyclizable substrate analogue GMPCPP[29]. In detail, the purified sGC was concentrated to $A_{280} = 6$ with an estimated concentration of 55.9 µM. For the haem-unliganded state sample, 5 mM $MgCl_2$, 0.5 mM fluorinated octyl-maltoside (FOM, Anatrace) were added; for the haem-oxidized state sample, 5 mM $MgCl_2$, 0.5 mM FOM, 100 µM NS2028 (Cayman Chemical), and 1 mM GTPγS (Sigma) were added; for the NO-activated state, 5 mM $MgCl_2$, 0.5 mM FOM, 1 mM noncyclizable substrate analogue GMPCPP (Biorbyt), and 1 mM DEA NONOate (Cayman Chemical) were added. For the β1(H105C) mutant sample, we added 5 mM $Mg^{2+}$ ions, 1 mM GMPCPP and 0.5 mM FOM to the protein. Protein samples were loaded onto glow-discharged Quantifoil 0.6/1 holey carbon gold grids and plunged into liquid ethane by Vitrobot Mark IV (Thermo Fisher Scientific).

**Disulfide bond cross-linking.** To generate the less-Cys sGC construct (sGC$^{LC}$), the cys-rich N-terminal 63 amino acids of α1 subunit were removed. Additional mutations of C176A, C239A, C669S, C455Y, and C460G were created in the α1 subunit and C292N in the β1 subunit. The coding sequences of α1$^{LC}$ with a C-terminal GFP-strep tag and β1$^{LC}$ without tags were cloned into modified BacMam expression vectors[32,33]. Then specific amino acids were mutated into cysteines using the Quick Change method. Cysteine mutants were transfected into HEK293F cells with polyethylenimine (PEI) (Polysciences) at a density of $2.0 \times 10^6$/ml. Cells were harvested 72 h after transfection and broken by passing through a syringe needle with 0.45 µm inner diameter ten times. Unbroken cells and large debris were removed by centrifugation at 14,800 r.p.m. for 10 min at 4 °C. sGC proteins were purified from supernatants using Streptactin Beads 4FF resin (Smart-Lifesciences, China). Protein samples were cross-linked on ice for 30 min by adding Cu(II) (1,10-phenanthroline)$_3$ to a final concentration of 30 µM to promote disulfide bond formation. Protein samples were subjected to 4–15% gradient SDS–PAGE (Beyotime Biotechnology, China) for separation either in non-reducing condition or reducing condition (in the presence of 100 mM DTT). The fluorescence was detected using a ChemiDoc MP (Bio-Rad) fluorescence imaging system.

**Cryo-EM data acquisition.** Cryo-grids were screened on a Talos Arctica electron microscope (Thermo Fisher Scientific) operating at 200 kV using a Ceta 16M camera (Thermo Fisher Scientific). The screened grids were transferred to a Titan Krios electron microscope (Thermo Fisher Scientific) operating at 300 kV with an energy filter set to a slit width of 20 eV. Images were recorded using a K2 Summit direct electron camera (Thermo Fisher Scientific) in super-resolution mode at a nominal magnification of 130,000×, corresponding to a calibrated super-resolution pixel size of 0.5225 Å. The defocus range was set from −1.5 µm to −2 µm. Each image was acquired as a 7.68-s movie stack (32 frames) with a dose rate of 6.25 e⁻Å⁻²s⁻¹, resulting in a total dose of about 48 e⁻Å⁻². All data acquisition was done using SerialEM.

**Cryo-EM data processing.** The data processing workflows are illustrated in Extended Data Figs. 1–4 and Extended Data Table 1. Super-resolution movie stacks were motion-corrected, mag-distortion corrected, dose-weighted, and binned to a pixel size of 1.045 Å by MotionCor2 1.1.0 using $5 \times 5$ patches[39]. Contrast transfer function (CTF) parameters were estimated from non-dose-weighted micrographs using Gctf v1.06[40]. Micrographs with ice or ethane contamination, empty carbon, and poor CTF fit (>5 Å) were manually removed. All classification and reconstruction was performed with Relion 3.0[41] unless otherwise stated. Particles were picked using Gautomatch (developed by Kai Zhang) and subjected to reference-free 2D classification to remove bad particles. Initial models were generated by cryoSPARC[42] using the selected particles from 2D classification. The particles

were further subjected to 3D classification to remove bad particles using the initial model, which was low-pass filtered to 30 Å as the reference. The particles selected from good 3D classes were re-centred and re-extracted, and their local CTF parameters were individually determined using Gctf v1.06[40]. These particles were imported into cisTEM[43] and subjected to 3D classification with auto-masking. The particles from the best 3D classes calculated by cisTEM were exported into Relion 3.0 and subjected to 3D auto-refinement to generate the consensus map. However, the two large lobes of sGC in the consensus maps showed blurry features, which were indicative of continuous conformational heterogeneities. Therefore, we divided the whole molecule into two bodies—the larger N lobe and the smaller C lobe—for further multibody refinement (Extended Data Figs. 1–4) in Relion 3.0, and the subsequent local map qualities were greatly improved (Extended Data Figs. 2a, b, 3e, f). In detail, two soft masks that cover the N lobe and C lobe were generated from the consensus map, which was edited manually in UCSF Chimera using the volume eraser tool[44]. 3D multi-body refinements[21] were performed using the two soft masks of the lobes and the parameters determined from previous 3D auto-refinement. The motions of the bodies were analysed by relion_flex_analyse in Relion 3.0. The two half-maps of each lobe generated by 3D multi-body refinement were subjected to post-processing in Relion 3.0. The masked and sharpened maps of each lobe were aligned to the consensus map using UCSF Chimera and summed to generate the composite map for visualization and interpretation. All of the resolution estimations were based on a Fourier shell correlation (FSC) of 0.143 cutoff after correction of the masking effect[45]. B-factors used for map sharpening were automatically estimated by the post-processing procedure in Relion 3.0.

**Model building.** The position of the $\beta 1$ H-NOX domain was first identified according to its distinguishable haem group density. Other domains were assigned by the domain–domain linkers that are visible in the post-processed map (Extended Data Fig. 2h). The homology models of individual H-NOX, PAS and catalytic domains were generated by the Phyre2 server[46] based on the structures of the human $\beta 1$ H-NOX domain (PDB: 5MNW), *M. sexta* $\alpha$ PAS domain (PDB: 4GJ4)[14] and human $\alpha 1$–$\beta 1$ catalytic domain heterodimer (PDB: 3UVJ)[16]. The models were placed into the corresponding composite maps using UCSF chimera[44] and manually rebuilt in Coot[47]. The composite maps were then converted into mtz files and the models were further refined by Phenix in reciprocal space[48] and Coot in real space. During model building, we found that the structures of the catalytic module in the haem-oxidized state and the haem-unliganded state were essentially the same, but we observed a positive difference density around the $\beta K$–$\alpha P$ loop of the $\alpha 1$ catalytic domain in the haem-oxidized state sample (Extended Data Fig. 2i). During the preparation of the haem-oxidized sample, we supplemented oxidizing reagent NS2028, substrate GTP$\gamma$S and cofactor $Mg^{2+}$ ions into the sGC protein. Therefore, based on the local chemical environment, this positive density might represent $Mg^{2+}$ ions together with highly negatively charged phosphate groups that possibly came from the decomposition of the GTP$\gamma$S molecule. However, these putative phosphate groups were not modelled. Volumes of the catalytic pocket were calculated using Caver[49] with the large probe radius 5 Å and the small probe radius 2.4 Å.

**Quantification and statistical analysis.** Global resolution estimations of cryo-EM density maps are based on the 0.143 FSC criterion[50]. The local resolution was estimated using Relion 3.0[41]. The number of technical replicates (*n*) and the relevant statistical parameters for each experiment (such as mean or standard deviation) are described in the figure legends. No statistical methods were used to pre-determine sample sizes.

**Reporting summary.** Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

## Data availability

Cryo-EM maps of the haem-unliganded, haem-oxidized, NO-activated and $\beta 1$(H105C) mutant sGC structures have been deposited in the EMDB under accession numbers EMD-9883, EMD-9884, EMD-9885 and EMD-9886, respectively. Atomic coordinates of the haem-unliganded, haem-oxidized and NO-activated sGC structures have been deposited in the PDB under accession numbers 6JT0, 6JT1 and 6JT2, respectively.

31. Vermehren-Schmaedick, A., Ainsley, J. A., Johnson, W. A., Davies, S. A. & Morton, D. B. Behavioral responses to hypoxia in *Drosophila* larvae are mediated by atypical soluble guanylyl cyclases. *Genetics* **186**, 183–196 (2010).
32. Goehring, A. et al. Screening and large-scale expression of membrane proteins in mammalian cells for structural studies. *Nat. Protoc.* **9**, 2574–2585 (2014).
33. Li, N. et al. Structure of a pancreatic ATP-sensitive potassium channel. *Cell* **168**, 101–110 (2017).
34. Kawate, T. & Gouaux, E. Fluorescence-detection size-exclusion chromatography for precrystallization screening of integral membrane proteins. *Structure* **14**, 673–681 (2006).
35. Tang, Q. et al. Structure of the receptor-activated human TRPC6 and TRPC3 ion channels. *Cell Res.* **28**, 746–755 (2018).
36. Olesen, S. P. et al. Characterization of NS 2028 as a specific inhibitor of soluble guanylyl cyclase. *Br. J. Pharmacol.* **123**, 299–309 (1998).
37. Brandwein, H. J., Lewicki, J. A., Waldman, S. A. & Murad, F. Effect of GTP analogues on purified soluble guanylate cyclase. *J. Biol. Chem.* **257**, 1309–1311 (1982).
38. Artz, J. D., Toader, V., Zavorin, S. I., Bennett, B. M. & Thatcher, G. R. In vitro activation of soluble guanylyl cyclase and nitric oxide release: a comparison of NO donors and NO mimetics. *Biochemistry* **40**, 9256–9264 (2001).
39. Zheng, S. Q. et al. MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat. Methods* **14**, 331–332 (2017).
40. Zhang, K. Gctf: Real-time CTF determination and correction. *J. Struct. Biol.* **193**, 1–12 (2016).
41. Zivanov, J. et al. New tools for automated high-resolution cryo-EM structure determination in RELION-3. *eLife* **7**, e42166 (2018).
42. Punjani, A., Rubinstein, J. L., Fleet, D. J. & Brubaker, M. A. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nat. Methods* **14**, 290–296 (2017).
43. Grant, T., Rohou, A. & Grigorieff, N. *cis*TEM, user-friendly software for single-particle image processing. *eLife* **7**, e35383 (2018).
44. Pettersen, E. F. et al. UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).
45. Chen, S. et al. High-resolution noise substitution to measure overfitting and validate resolution in 3D structure determination by single particle electron cryomicroscopy. *Ultramicroscopy* **135**, 24–35 (2013).
46. Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N. & Sternberg, M. J. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protoc.* **10**, 845–858 (2015).
47. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D* **66**, 486–501 (2010).
48. Adams, P. D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D* **66**, 213–221 (2010).
49. Jurcik, A. et al. CAVER Analyst 2.0: analysis and visualization of channels and tunnels in protein structures and molecular dynamics trajectories. *Bioinformatics* **34**, 3586–3588 (2018).
50. Rosenthal, P. B. & Henderson, R. Optimal determination of particle orientation, absolute hand, and contrast loss in single-particle electron cryomicroscopy. *J. Mol. Biol.* **333**, 721–745 (2003).

**Author contributions** L.C. initiated the project and screened expression constructs. R.L. purified protein and prepared the cryo-EM sample. Y.K. and R.L. collected the cryo-EM data with the help of J.-X.W. Y.K. processed the cryo-EM data with the help of L.C. L.C. built and refined the atomic model. Y.K. did the enzymatic activity assay. R.L. did the disulfide bond cross-linking experiment. All authors contributed to manuscript preparation.

**Competing interests** The authors declare no competing interests.
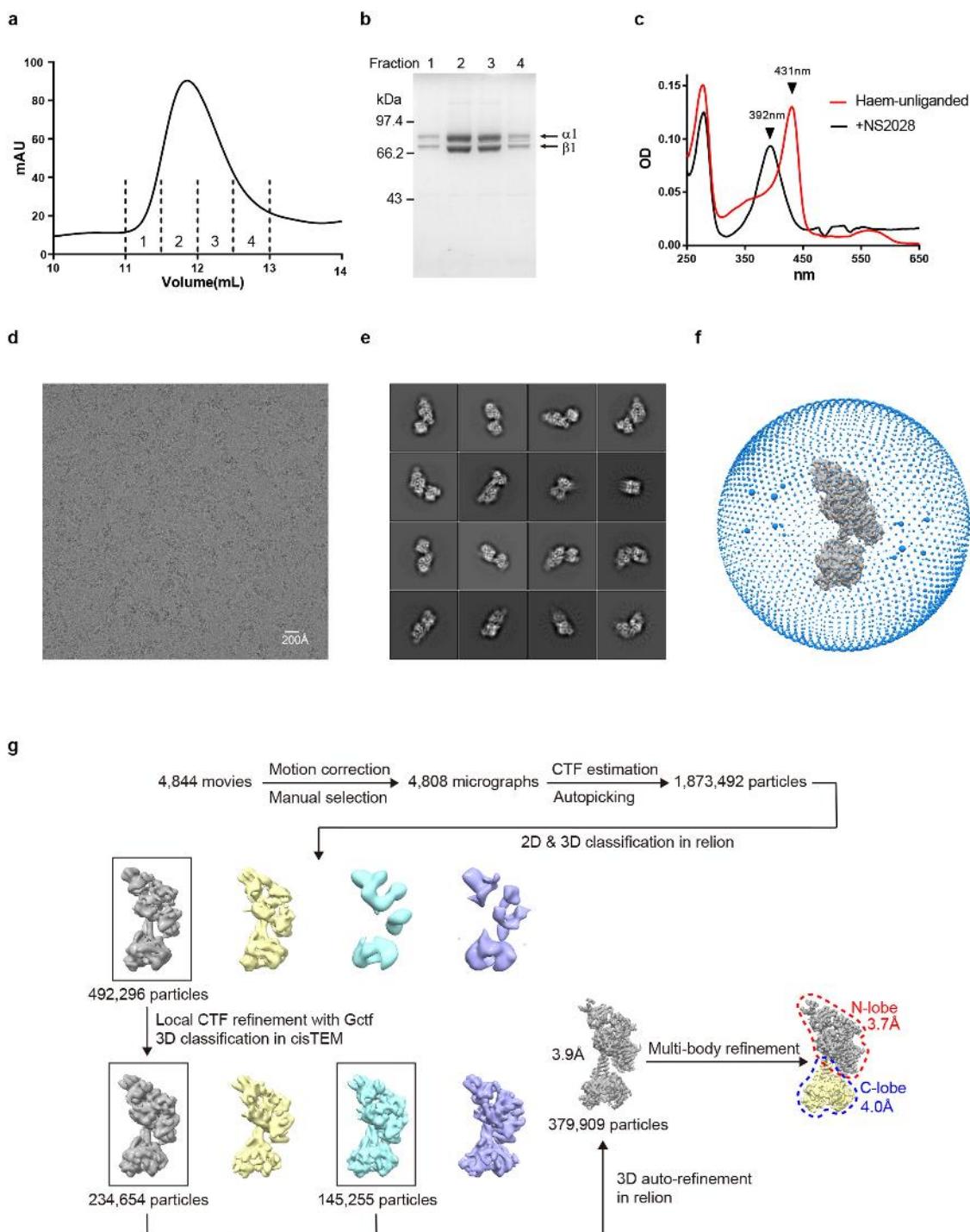
**Extended Data Fig. 1 | Biochemical characterization of the human α1β1 sGC heterodimer protein and single-particle cryo-EM data processing procedure for sGC in the inactive (haem-oxidized) state. a**, Size-exclusion chromatography of sGC on a superdex 200 column. The fractions indicated by dashed lines were pooled for cryo-EM sample preparation. **b**, SDS–PAGE of the size-exclusion chromatography fractions labelled in **a**. Arrows show the positions of the α1 and β1 subunits. For gel source data, see Supplementary Fig. 1. **c**, UV-vis spectra of purified sGC before (red) and after (black) treatment with the haem oxidant NS2028.

The positions of the Soret peaks are indicated by arrowheads. **a–c**, The experiments were repeated independently three times with similar results. **d**, A representative raw micrograph of sGC in the inactive (haem-oxidized) state. **e**, Representative 2D class averages of sGC in the inactive (haem-oxidized) state. **f**, The angular distribution for the consensus refinement of the inactive (haem-oxidized) state is indicated by the sizes of spheres. **g**, The cryo-EM data processing workflow for sGC in the inactive (haem-oxidized) state.
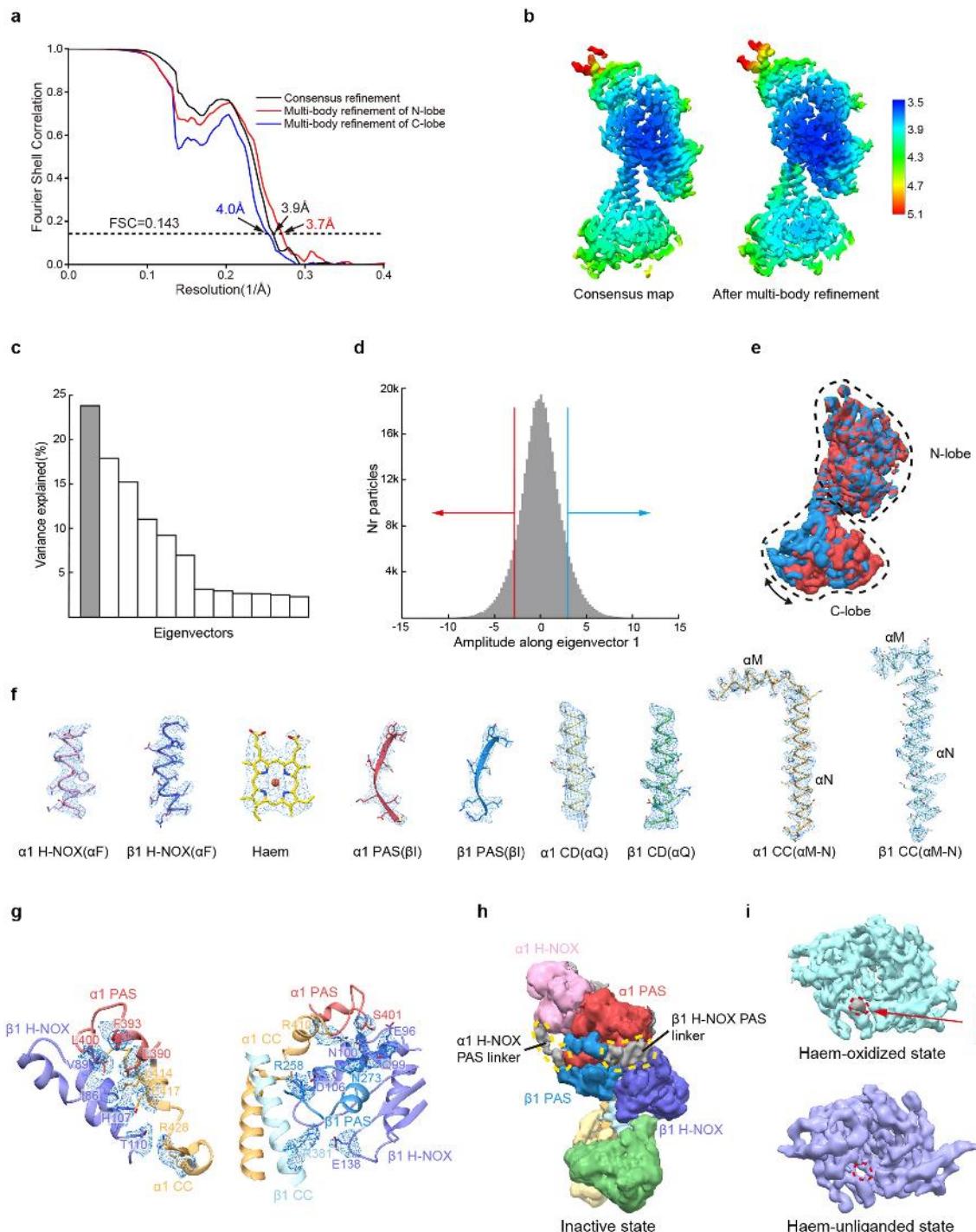
**Extended Data Fig. 2 | Conformational heterogeneity and local density quality of sGC in the inactive (haem-oxidized) state. a**, Gold-standard FSC curves of haem-oxidized sGC after correction for masking effects. Resolution estimations were based on the criterion of the FSC 0.143 cutoff. **b**, Local resolution distribution of the composite map of sGC in the inactive (haem-oxidized) state. **c**, Histogram of the eigenvectors that contribute to the variance. The top eigenvector is highlighted in grey. **d**, Histogram of the amplitudes along the top eigenvectors shows monomodal distribution. Particle populations with amplitudes less than −3 or greater than 3 are indicated as red and blue arrows, respectively. **e**, A 4 Å low-pass filtered map reconstructed from particles indicated

as red and blue arrows in **d**. N-lobes were used for alignment. **f**, Representative cryo-EM densities of fragments from each individual domain. **g**, Representative cryo-EM densities of several key residues involving the interactions between β1 H-NOX and adjacent domains in the inactive (haem-oxidized) state. **h**, The cryo-EM map of the sGC in the haem-oxidized state. The putative linkers between the H-NOX and PAS domains are shown in grey. The B-factor of the map was adjusted to −100 Å² during the post-processing procedure to visualize features with high flexibility. **i**, Cryo-EM maps of the catalytic module in the haem-oxidized state (cyan) and the haem-unliganded state (purple). The density of the putative phosphate groups is shown in grey.

a



b



c



d



e



f



Consensus map     After multi-body refinement
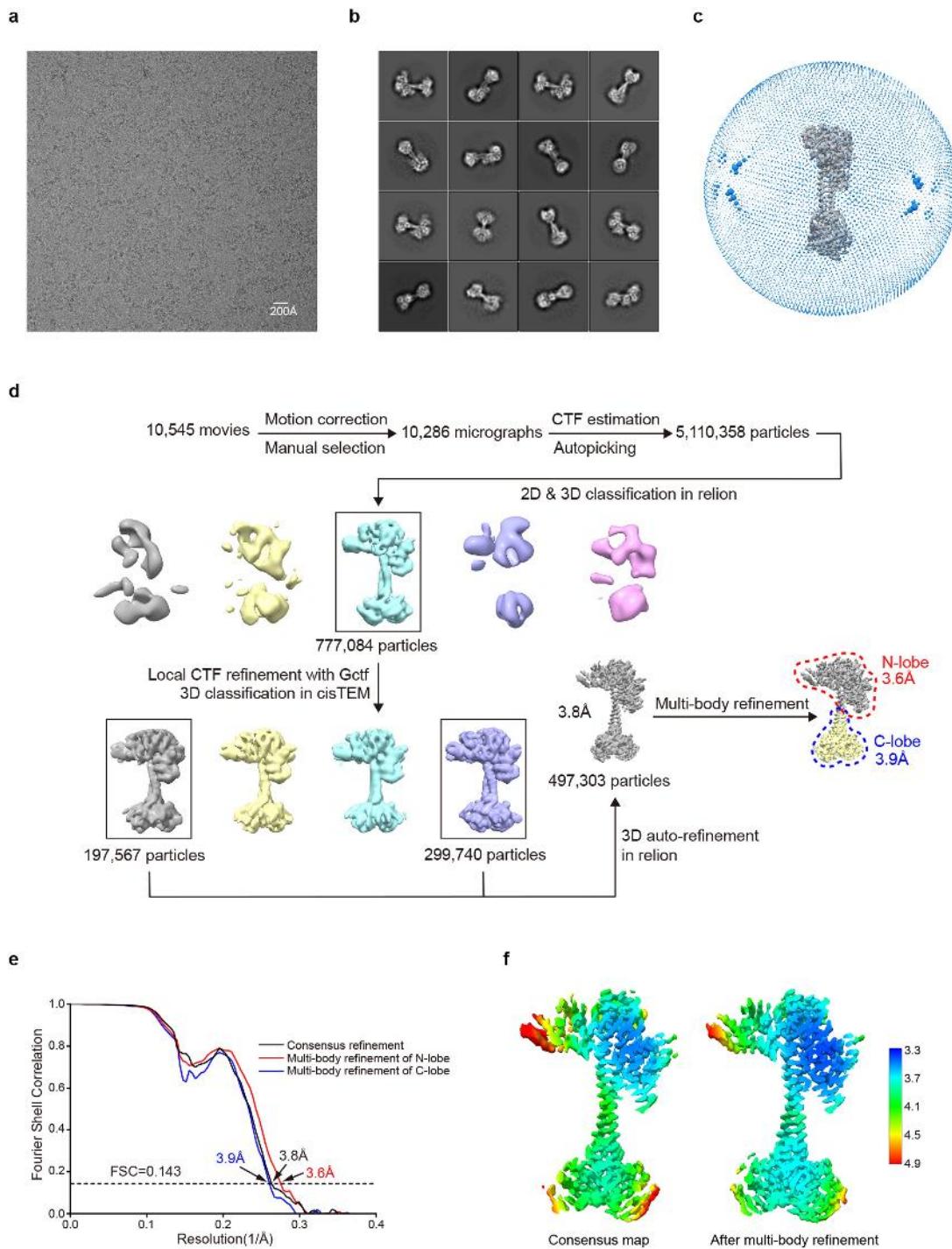
**Extended Data Fig. 3 | Single particle cryo-EM data processing of the sGC sample in the NO-activated state. a,** Representative raw micrograph of sGC in the NO-activated state. **b,** Representative 2D class averages of sGC in the NO-activated state. **c,** The angular distribution for the consensus refinement of the NO-activated state is indicated by the sizes of the spheres. **d,** The cryo-EM data processing workflow for sGC in the NO-activated state. **e,** Gold-standard FSC curves (after correction for the masking effects) of NO-activated sGC. Resolution estimations are based on the criterion of FSC 0.143 cutoff. **f,** Local resolution distribution of the composite map of sGC in the NO-activated state.
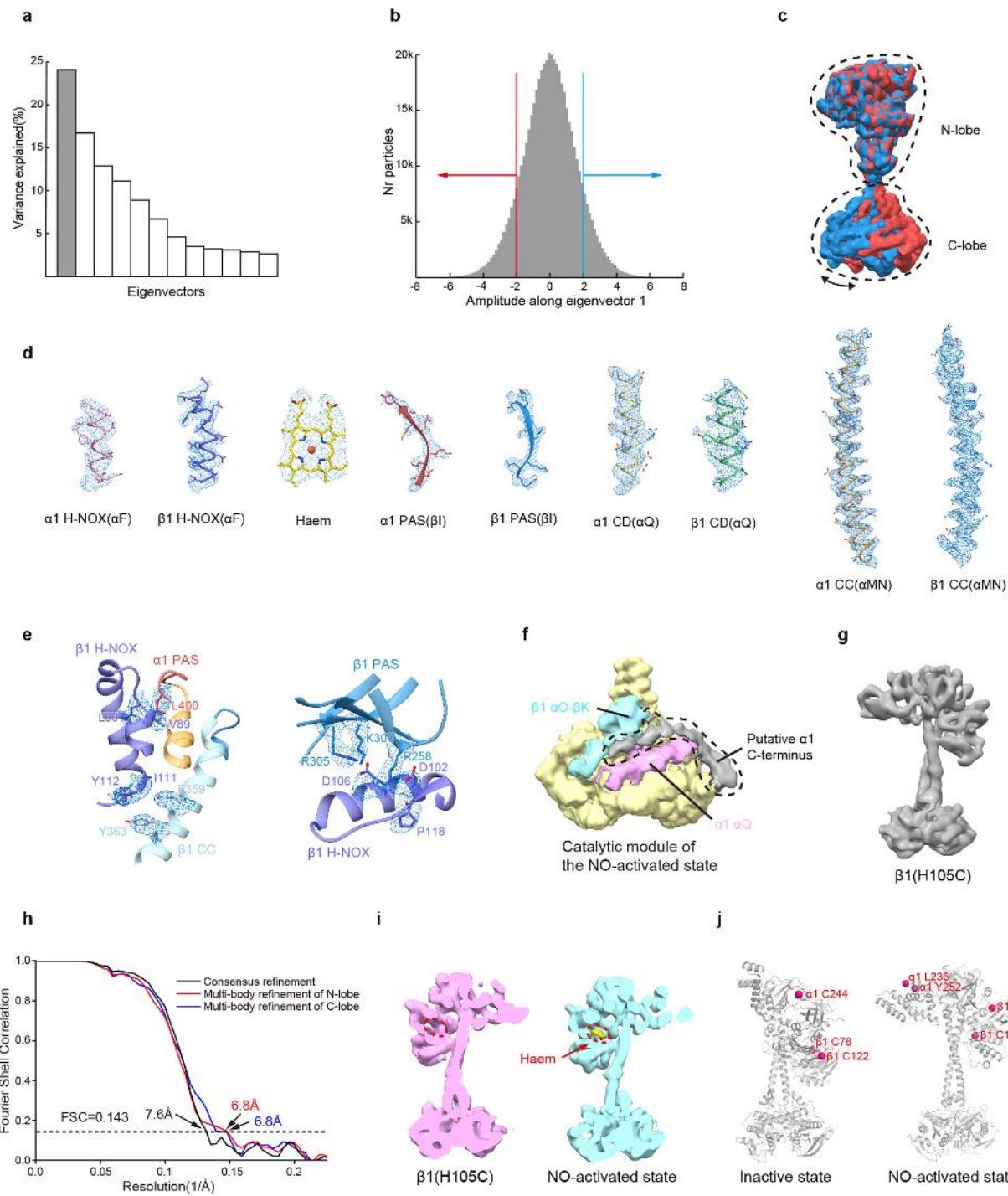
**Extended Data Fig. 4 | Conformational heterogeneity and local density quality of sGC in the NO-activated state. a**, Histogram of the eigenvectors that contribute to the variance. The top eigenvector is highlighted in grey. **b**, Histogram of the amplitudes along the top eigenvectors shows monomodal distribution. Particle populations with amplitudes less than −2 or greater than 2 are indicated with red and blue arrows, respectively. **c**, 4 Å low-pass filtered maps reconstructed from particles indicated as red and blue arrows in **b**. N-lobes were used for alignment. **d**, Representative cryo-EM densities of fragments from each individual domain. **e**, Representative cryo-EM densities of several key residues involving the interactions between β1 H-NOX and adjacent domains in the NO-activated state. **f**, Cryo-EM map of the catalytic module in the NO-activated state (yellow). The putative density of the α1 C terminus is shown in grey, αQ of the α1 subunit is shown in pink, and

the αO–βK fragment of the β1 subunit is shown in cyan. The B-factor of the map was adjusted to −100 Å² in post-processing to visualize features with high flexibility. **g**, The side view of the cryo-EM map of the β1(H105C) mutant sGC. **h**, Gold-standard FSC curves (after correction for masking effects) of the β1(H105C) mutant sGC. Resolution estimations were based on the criterion of FSC 0.143 cutoff. **i**, Cryo-EM map of β1(H105C) mutant sGC (pink) and NO-activated sGC (cyan). The haem density is shown in yellow. The map of NO-activated sGC was low-pass filtered to 6.8 Å. **j**, The locations of cysteine residues that are involved in sGC desensitization (α1 C244, β1 C78 and β1 C122) are indicated with their Cα atoms shown as red spheres. Because the loop containing α1 C244 is disordered in the NO-activated state, only the termini of the loop (α1 L235 and α1 Y252) containing α1 C244 are labelled.

```
                              αA                    αB                                                   αC
Hs_sGCα1    1  ---------MFCTKLKDLKITG-----ECPFS--LLAPGQVPNESSEEAAGSSESCKATVPICQDIPEKNIQESLPQRKT  64
Hs_sGCα2    1  MSRRKISSESFSSLGSDYLETSPEEEGECPLSRLCWNGSRSPPGPLEPSPAAAAAAAAPAPTPAASAAAAAATAGARRVQ  80
Dr_sGCα1    1  ---------MFCAKLKDLKIAG-----ECPFS--MVLG----NEEPADFEGSN--AVNVVPVSREVQG-NITNILPRRKI  57
Dr_sGCα2    1  ----------------LVKYFRR-------------WHG----------------------------------------RLK  13
Ms_sGCα1    1  ---------MTCPFRRASSQHQ------FANGG---------------------SSAPKKPEFRSRTSSVHLTGPEEEDG  44

                   αA                    αB                                                   αC
Hs_sGCα1   65  SRSRVYLHTLAESICKLIFPEFERLNVALQRTLAKHKIK--------ESRKSLE-------------REDFEKTIAEQAV  123
Hs_sGCα2   81  RRRRVVNLDSLGESISRLTAPSPQTIQQTLKRTLQYYEHQVIGYRDAEKNFHNISNRCSYADHSNKEEIEDVSGILQCTAN  160
Dr_sGCα1   58  SRSKVNLHSLGESVRKLVCPEFEKLRNALIRVIKQQIDTSCPVACSSDVSHTLDN-----------AEQLSDVMKNYSL  125
Dr_sGCα2   14  ARR---------SLVRVFAG------------------------------------------------------PVGIS  29
Ms_sGCα1   45  ERNTLTLKHMSEALQLLTAPSNECLHAAVTSLTKNQSDHYHKYNCLRRLPDDVKTCRN---------YAYLQEIYDAVRA  115

                   αD                    αE            αF                           βA
Hs_sGCα1  124  AAGVPVEVIKESLGEEVFKICYEED---ENILGVVGGTLKDFLNSFSTLLKQSSHCQEAGKRGRLEDAS-ILCLDKEDD-  198
Hs_sGCα2  161  ILGLKFEEIQKRFGEEFFNICFHEN---ERVLRAVGGTLQDFFNGFDALEHIR--TSFGKQATLESPS-FLCKELPEG-  233
Dr_sGCα1  126  KTGIPMDALKISLGLEVFRDCYEED---GHILRVVGGALHDFLNSFNVLLKQSTPPTPEIQR-YIHQAS-ILCLDKDPG-  199
Dr_sGCα2   30  VWSVDFAAIRGRFGEEFFGLCLEEN---ERVLRAIGGNLQDFFNGFDALEHIR--TSCGRRASSEAPS-FLCKDAPQEN  103
Ms_sGCα1  116  TDSVNTKDFMAKLGEYLILTAFSHNCRLERAFKCLGTNLTEFLTTLDSVHDVLHDQDTPLKDETMEYEANFVCTTSQEG-  194

                   βB               αG           βC       αH
Hs_sGCα1  198  ----------FLHVYYFPPKRTTSLILPGIIKAAAHVLYETEVEVSLMPPC----------FHNDCSEFVNQP------  251
Hs_sGCα2  233  ----------TLMLHYFHPHHIVGFAMLGMIKAAGKKIYRLDVEVEQV-------ANEX--LCSDVSNPGNCS------  287
Dr_sGCα1  199  ----------LLTVYFFNPLPSTELFFSGIIQAAACLLYCTRVEVRMDVGG----------KDSGALQADHQP------  252
Dr_sGCα2  104  EETNAESGAGRTLILHCFNPDPTVGVAMPGLIRAAARRIYQSEVSVEEAPPTWLHTANEDGELSADSSPSSSPSPPSSSA  183
Ms_sGCα1  194  ----------KIQLHLTTESEPVAYLLVGSLKAIAKRLYDTQTDIRLRSYTN---DPRRFRYEINAVPLHQKS------  254

                   βD                     αI       βE       βF      αJ
Hs_sGCα1  251  -----YLLYSVHMKSTKPSLSPS--KPQSSVIPTSLFCKTFPFHFMFDKDMTILQFGNGIRRLMNR-RDFQGKPNPEEY  323
Hs_sGCα2  287  --CLTFLIKECENTNIMKNLPQGTSQVPADLRISINFECRAFPFHLMFDPSMSVLQLGEGLRRFQLRC--DTHKVLKPEDC  363
Dr_sGCα1  252  -----HLLYSVVVRDGR-SLTPSPMRTHTSGDIPLSLLYSTFPFHILLDQEMGLLQIGDGLRRRLGRCRDGQRRPAFNEH  326
Dr_sGCα2  184  PACLSPLIREVHTPPTSSSRSRQLSRSPLDVRIGLSTFCRAFPFHLVLGSNMEVLQVGEGLRKQAKT--DPHRTLIFSNS  261
Ms_sGCα1  254  --------KEDSCELVNEAASVATSTKVTDEKIGVASFCKAFPWHFITDKRLELVQLGAGFMRLFGTHLATHG-SSLGTY  325

                   βG          αK        βH                   βI         βJ        αL
Hs_sGCα1  324  FEILTPK-INQTFSGIMTMLNMQFVVRVRRWDNSVKKSS-RVMDLKGQMIYIVESSAILFLGSPCVDRLEDFTGRGLYLS  401
Hs_sGCα2  364  FEIVSPK-VNATFERVLLRLSTPFVIRTKPEASGSENKD-KVMEVKGQMIHVPESNSILFLGSPCVDKLDELMGRGLHLS  441
Dr_sGCα1  327  FAIVSPE-IRASFQDILTMLNTQFLLRVKQHGASSADSPCKHMDLKGQMIFMSEMSALLFLGSPCVDKLEELTGRCLYLS  405
Dr_sGCα2  262  FELVSPR-IPCSFQNILRLSTPFTIRTRPDGSSLDCRE-KVMELKGQMLHLPESNSIMFLGSPRVDRLEELMGRGLHLS  339
Ms_sGCα1  326  FRLLRPRGVPLDFREILKRVNTPFMFCLKMPGSTALAEG---LEIKGQMVFCAESDSLLFVGSPFLDGLEGLTGRGLFIS  402

                      αM              αN                    αO       βK
Hs_sGCα1  402  DIPIHNALRDVVLIGEQARAQDGSLKKRLGKLKATLEQAHQALEEEKKKTVDLLCSIFPCEVAQQLWQGQVVQAKKFSNVT  481
Hs_sGCα2  442  DIPIHDATRDVILVGEQAKAQDGSLKKRMDKLKATLERTHQALEEEKKKTVDLLYSIFPGDVAQQLWQGQVQARKFDDVT  521
Dr_sGCα1  406  DIPIHNALRDVVLVCEQTKAQDGSLKKRLGKAALEQAHQALEEEKKRRIVELLFTIFPGNVAQRLWQGLPVQAKKFDHVT  485
Dr_sGCα2  340  DIPIHDATRDVILVGEQAKAQDGSLKKRMDKLKALTLEKTHQALEEEKKRRTVDLLYSIFPGDVAQQLWQGLPVQAKKFDDVT  419
Ms_sGCα1  403  DIPLHDATRDVILVGEQARAQDGSLRRRMDKLKNSIEEASKAVDKEREKNVSLLHLIFPPHIAKRLWLGEKIEAKSHDDVT  482

                      αP               αQ            βL    βM                   αR
Hs_sGCα1  482  MLFSDIVGFTAICSQCSPLQVITMLNALYTRFDQQCGELDVYKVETIGDAYCVAGGLHKESDTHAVQIALMALKMMELSD  561
Hs_sGCα2  522  MLFSDIVGFTAICAQCTPMQVISMLNELYTRFDHQCGFLDIYKVETIGDAYCVAAGLHRKSLCHAKPIALMALKMMELSE  601
Dr_sGCα1  486  VLFSDIVGFTAICSHCSPMQVVNMLSELYTRFDHHCGELDVYKVETIGDAYCVAGGLHKESPTHAVQIALMALKMMELSD  565
Dr_sGCα2  420  MLFSDIVGFTAVCAQCTPMQVISMLNELYTRFDYQCGILDVYKIETIGDAYCVASGLHRKIDSHAKPIALMALKMMELSE  499
Ms_sGCα1  483  MLFSDIVGFTSICATATPPMMVIAMLEDLYSVFDIPCGEILDVYKVETIGDAYCVASGLHRKVETHAPQIAWMALRMVETCA  562

                      βN              βO       αS       βP   αT       βQ
Hs_sGCα1  562  EVMSPHGEPIKMRIGLHSGSVFAGVVGVKMPRYCLFGNNVTLANKFESCSVPRKINVSPTTYRLLKDCPGFVFTPRSREE  641
Hs_sGCα2  602  EVLTPDGRPIQMRIGIHSGSVLAGVVGVRMPRYCLFGNNVTLASKFESGSHPRRINVSPTTYQLLKRESFTEFPRSREE  681
Dr_sGCα1  566  EVITPMGEVIRMRIGIHSGSVLAGVVGVKMPRYCLFGNNVTLASKFESCSLPRKINVSPTTYRLLKDCPEFTLIPRTRED  645
Dr_sGCα2  500  EVLTPDGKPIKLRIGIHSGSVLAGVVGVMMPRYCLFGNNVTLASKFESGSHPRCINVSPTTYQLLRDDRSFTEFPRSRQE  579
Ms_sGCα1  563  QHLTHFGNPIKMRIGLHTGTVLAGVVGKTMLKYCLFGHNVTLANKFESGSEPLKINVSPTTYEWLIKFPGFDMEPRDRSC  642

                      βR
Hs_sGCα1  642  LPPNFPSEIPGICHFLDAYQ-QGTN----SKPCFQKKDVEDGNANFLGKASGID---  690
Hs_sGCα2  682  LPDNFPKEIPGICYFLEVR--TGPKPPKPSLSSSRIKKVSYNIGTMFLRETSL----  732
Dr_sGCα1  646  LPPNFPTDIPGVCYFLQAHD-QKTN----AAP-------FETQP--------SG----  679
Dr_sGCα2  580  LPDNFPKEIPGICYFLEAGK-SQSHASLTSTRSAPIRKVSYNIGTMFLRETSL----  631
Ms_sGCα1  643  LPNSFPKDIHGTCYFLEKYTHPGTDPGEPQVKHIREALKDYGIGQANSTDVDTEEPT  699
```

**Extended Data Fig. 5 | Sequence alignment of the sGC α-subunit.** The sequences of the *Homo sapiens* α1 subunit, *H. sapiens* α2 subunit, *Danio rerio* α1 subunit, *D. rerio* α2 subunit, and *M. sexta* α1 subunit are aligned. Conserved residues are coloured in grey. Residues that are mutated to cysteines for oxidative cross-linking are indicated with a black box. Mutations for activity assay are indicated with a red box. The residues corresponding to H105 in the β1 subunit are indicated with a yellow box. Secondary structural elements are indicated as follows: arrows, β-sheets; cylinders, α-helices; lines, loops. Unmodelled residues are shown as dashed lines. The colours of arrows and cylinders are as in Fig. 1a.
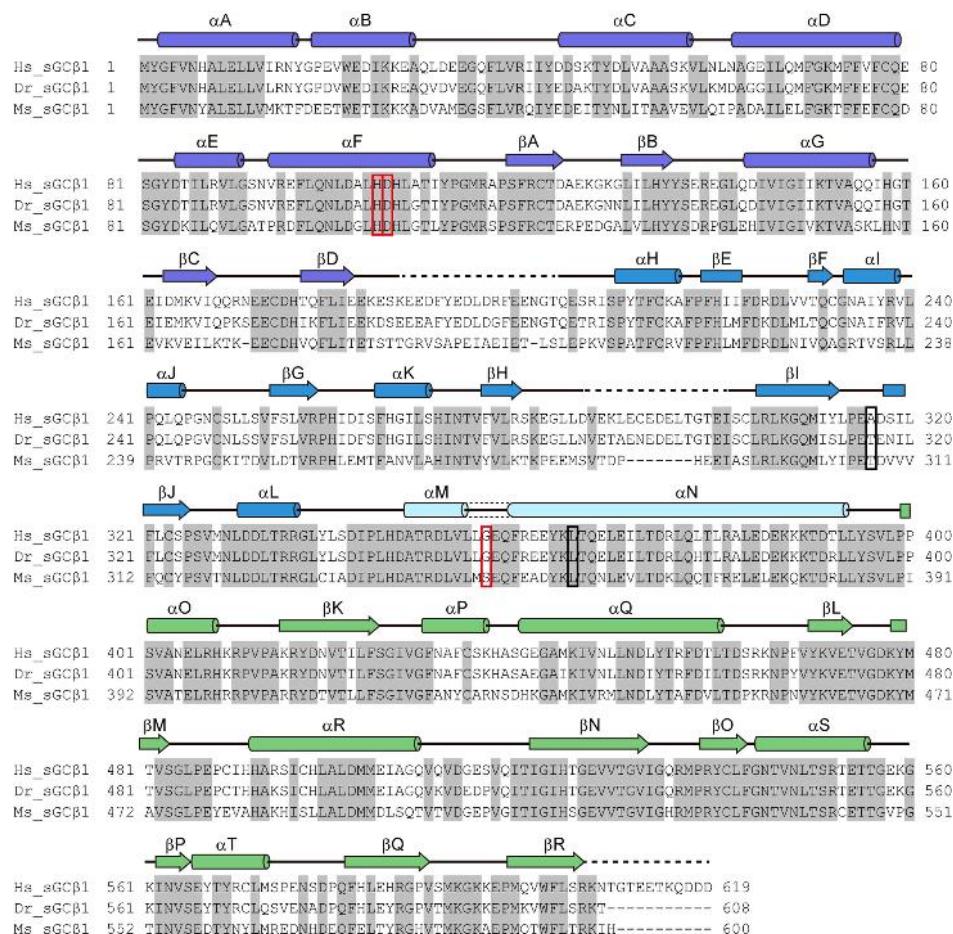
**Extended Data Fig. 6 | Sequence alignment of the sGC β1 subunit.** The sequences of the *H. sapiens* β1 subunit, *D. rerio* β1 subunit, and *M. sexta* β1 subunit are aligned. Conserved residues are coloured in grey. Residues that are mutated to cysteines for oxidative cross-linking are indicated with a black box. Mutations for activity assay are indicated with a red box. Secondary structural elements are indicated as follows: arrows, β-sheets; cyinders, α-helices; lines, loops. Unmodelled residues are shown as dashed lines. The colours of arrows and cylinders are as in Fig. 1a.
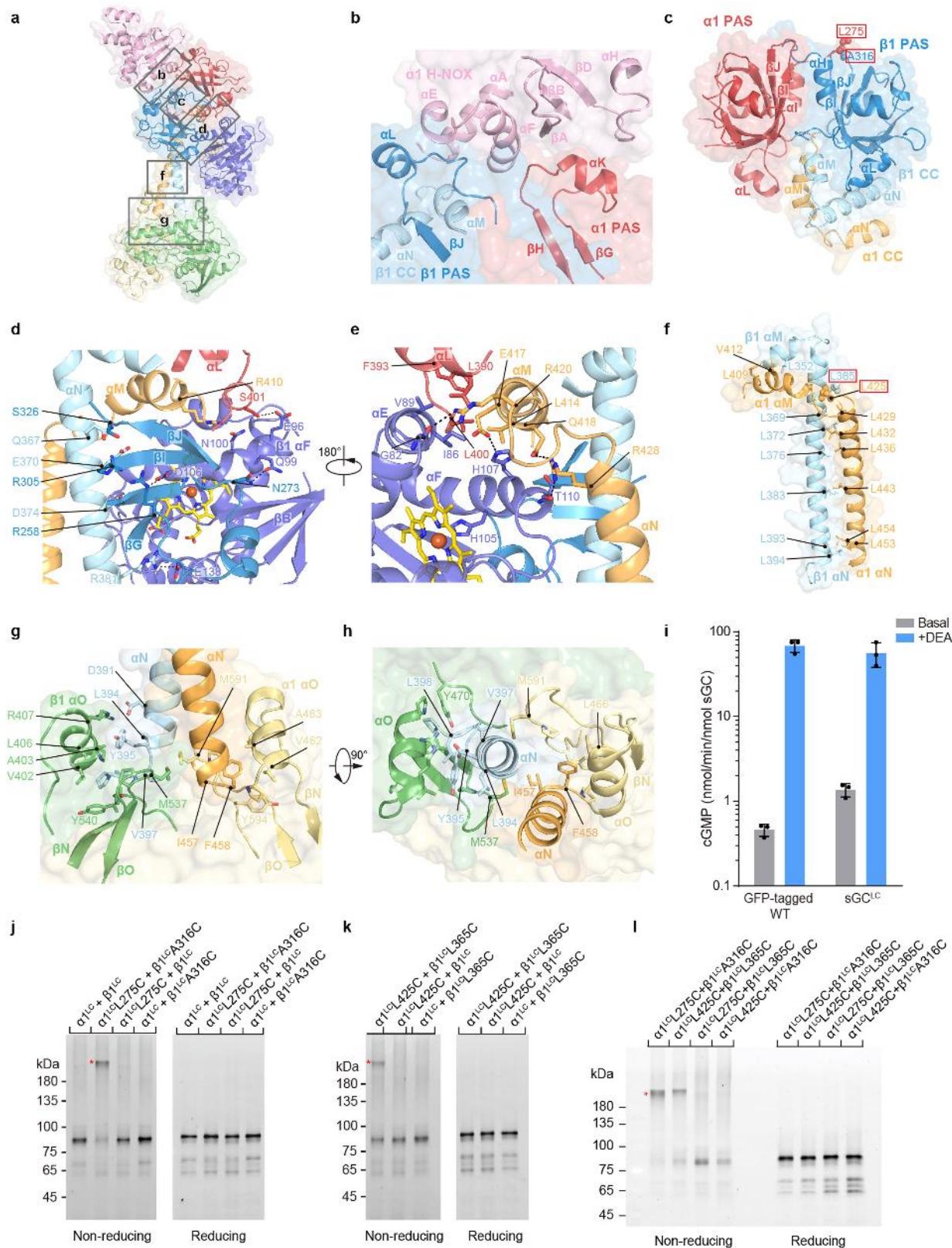
**Extended Data Fig. 7 |** See next page for caption.

**Extended Data Fig. 7 | Domain–domain interfaces of sGC in the inactive state. a**, Side view of soluble guanylate cyclase in the inactive state, highlighting key interfaces (grey rectangles). Each domain is coloured as in Fig. 1a. The surface of sGC is shown in transparency. **b**, The interface between the α1 H-NOX domain and the PAS domains boxed in **a**. **c**, The interface between the PAS domains boxed in **a**. **d**, The interface between β1 H-NOX and adjacent domains boxed in **a**. **e**, A 180° rotated view compared to **d**. **f**, The structure of the transducer module boxed in **a**. The side chains of α1 L425 and β1 L365 that are in close proximity are shown as spheres. **g**, The interface between the transducer module and the catalytic module boxed in **a**. **h**, A 90° rotated top view compared to **g**. **i**, End-point activity of the less-Cys construct (sGC$^{LC}$, α1$^{LC}$ + β1$^{LC}$) compared to the wild-type sGC with CGFP. Mean ± s.d., $n = 3$ biologically independent samples. **j**, SDS–PAGE of the in vitro disulfide bond cross-linking experiment of α1$^{LC}$ (L275C) with β1$^{LC}$(A316C) mutants under reducing and non-reducing conditions. The in-gel GFP fluorescence of the α1 subunit is shown in black on a white background. The position of cross-linked heterodimer is indicated with a red asterisk.

Oxidative cross-linking happened only when the cysteine mutants, α1(L275C) and β1(A316C), were present in both subunits simultaneously. The experiments were repeated independently three times with similar results. For gel source data, see Supplementary Fig. 1. **k**, SDS–PAGE of the in vitro disulfide bond cross-linking experiment of α1$^{LC}$(L425C) with β1$^{LC}$(L365C) under reducing and non-reducing conditions. Oxidative cross-linking happened only when the cysteine mutants, α1(L425C) and β1(L365C), were present in both subunit simultaneously. For gel source data, see Supplementary Fig. 1. The experiments were repeated independently three times with similar results. **l**, SDS–PAGE of the in vitro disulfide bond cross-linking experiment of α1$^{LC}$(L275C) with β1$^{LC}$(L365C) and α1$^{LC}$(L425C) with β1$^{LC}$(A316C) under reducing and non-reducing conditions. In contrast to α1(L275C) with β1(A316C) and α1(L425C) with β1(L365C), α1(L275C) did not crosslink with β1(L365C), and α1(L425C) did not crosslink with β1(A316C), owing to their long spatial distance. For gel sourcfe data, see Supplementary Fig. 1. The experiments were repeated independently twice with similar results.
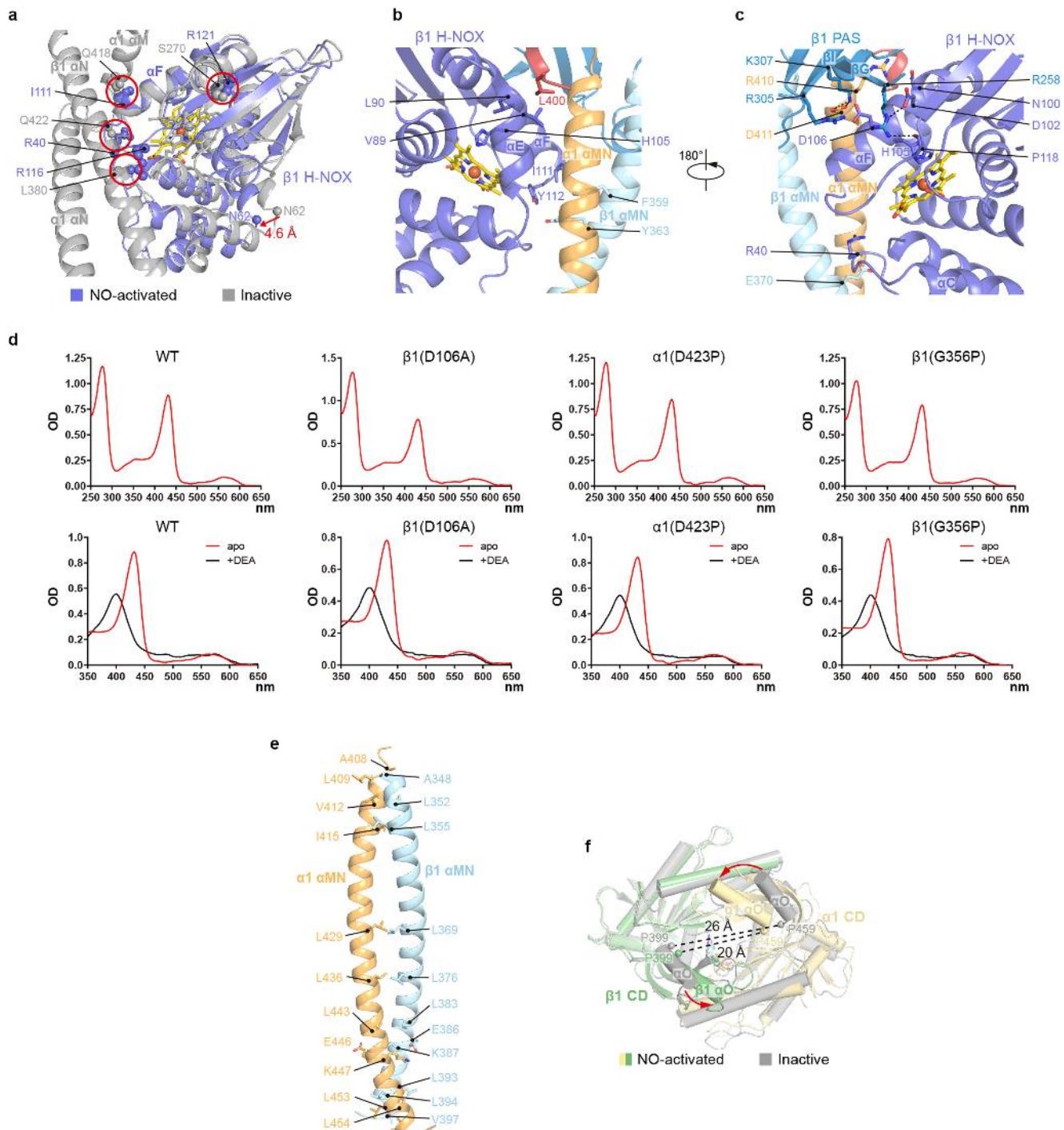
**Extended Data Fig. 8 | Domain–domain interfaces of sGC in the NO-activated state. a**, Superposition of the NO-bound β1 H-NOX domain structure (purple) onto the inactive state structure (grey) by alignment of the αF helices. The steric clashes between the side chains of the NO-bound β1 H-NOX domain (purple sphere) and the side chains of the PAS and CC domains of the inactive state (grey sphere) are marked by red circles if their atom-to-atom distances are smaller than 2.2 Å. The arrow indicates the positional change of the Cα atoms of β1 N62 induced by NO binding. **b**, The interface between β1 H-NOX and adjacent domains of sGC in

the NO-activated state. **c**, A 180° rotated view compared to **b**. **d**, Soret peaks of the sGC mutants show markedly decreased NO activation. The experiments were repeated independently twice with similar results. **e**, The transducer module in the NO-activated state, coloured as in Fig. 1a. **f**, Top view of the structural comparison of the catalytic module between the inactive state (grey) and the NO-activated state (coloured). The GMPCPP molecule is shown as sticks. The Cα atoms of α1 P459 and β1 P399 are shown as spheres.

**a**
- Haem-oxidized sGC
- Haem-unliganded sGC

**b**
N-terminal subdomain
C-terminal subdomain
α1-αA
- α1 H-NOX
- β1 H-NOX

**c**
N
C
- α1 PAS
- Ms PAS α domain (PDB: 4GJ4)

**d**
N
C
- β1 PAS
- Ms PAS α domain (PDB: 4GJ4)

**e**
β1 CD
α1 CD
- Inactive CD
- Isolated sGC CD (PDB: 4NI2)

**f**
90°
13°
αQ
- Inactive CD
- Isolated sGC CD (PDB: 4NI2)

**g**
N431
T602
D477
L596
C595
α1 CD
β1 CD
- Inactive CD
- Active adenylate cyclase CD (PDB: 1CJU)

**h**
αF
- NO-activated β1 H-NOX
- Inactive β1 H-NOX

**i**
β1 CD
α1 CD
- NO-activated CD
- Isolated sGC CD (PDB: 4NI2)

**j**
90°
26°
αQ
- NO-activated CD
- Isolated sGC CD (PDB: 4NI2)

**k**
Pocket
Pocket
GMPCPP
-61.970   +61.970
Inactive state
-68.446   +68.446
NO-activated state

**l**
α1 CD
β1 CD
GMPCPP / ddATP
- NO-activated
- Active adenylate cyclase (PDB: 1CJU)

**m**
β1 PAS
α1 PAS
- NO-activated PAS dimer
- Inactive PAS dimer

**Extended Data Fig. 9 |** See next page for caption.

**Extended Data Fig. 9 | Structural comparisons of each domain.**
**a**, Structural comparison of the full-length human sGC between the haem-oxidized state (coloured) and the haem-unliganded state (grey). **b**, Structural comparison between the α1 H-NOX domain (pink) and the β1 H-NOX domain (blue) in the inactive state. Both α1 H-NOX and β1 H-NOX share common structural features with prokaryotic H-NOX domains, which are composed of both N-terminal and C-terminal subdomains. The N-terminal αA helix of the α1 subunit that occupies the haem binding pocket is shown in red. The haem molecule of the β1 H-NOX domain is shown as a yellow stick. The approximate boundaries of N-terminal and C-terminal subdomains are indicated by dashed lines. **c**, Structural comparison between the human α1 PAS domain (red) and the *M. sexta* Ms α PAS domain (grey, PDB ID:4GJ4). **d**, Structural comparison between the human β1 PAS domain (blue) and the *M. sexta* Ms α PAS domain (grey, PDB ID:4GJ4). **e**, Structural comparison between the catalytic module of the full-length sGC in the inactive state (coloured) and the isolated catalytic domain heterodimer (grey, PDB ID: 4NI2). The β1 subunit was used for structural alignment. **f**, 90° rotated view compared to **e**. **g**, Structural comparison between the catalytic module of the full-length sGC in the inactive state (coloured) and the catalytic domain of the active adenylate cyclase (grey, PDB ID: 1CJU, chain A&B). The β1 subunit was used for structural alignment. The residues of sGC that are within 2.2 Å of the substrate are considered as sterical clashes and shown as red spheres. **h**, Structural comparison between the NO-activated state (purple) and the inactive state (grey) of the human β1 H-NOX domain. The N-terminal subdomain was used for alignment and the movements are indicated as red arrows. **i**, Structural comparison between the catalytic module of the full-length sGC in the NO-activated (colored) and the isolated catalytic domain heterodimer (grey, PDB ID: 4NI2). The β1 subunit was used for structural alignment. A inter-domain rotational conformational change is observed. **j**, A 90° rotated view compared to **i**. **k**, Cutaway views of the sGC catalytic module in the inactive state and the NO-activated state. The catalytic module is shown in surface representation colored by electrostatic potential calculated in Pymol. The pockets inside the catalytic module are indicated by arrows. GMPCPP molecule is shown as sticks. **l**, Structural comparison of the catalytic core between the active adenylate cyclase (grey, PDB ID: 1CJU, chain A and B) and sGC in the NO-activated state (coloured). **m**, Structural comparison between the NO-activated state (coloured) and the inactive state (grey) of the human α1 and β1 PAS heterodimer.

**Extended Data Table 1 | Cryo-EM data collection, refinement and validation statistics**

| | Haem-unliganded state 6JT0 EMD-9883 | Haem-oxidized state 6JT1 EMD-9884 | NO-activated state 6JT2 EMD-9885 | β1(H105C) EMD-9886 |
|---|---|---|---|---|
| **Data collection and processing** | | | | |
| Magnification | 130,000 × | 130,000 × | 130,000 × | 130,000 × |
| Voltage (kV) | 300 | 300 | 300 | 300 |
| Electron exposure (e–/Å$^2$) | 48 | 48 | 48 | 48 |
| Defocus range (μm) | -1.5 to -2.0 | -1.5 to -2.0 | -1.5 to -2.0 | -1.5 to -2.0 |
| Pixel size (Å) | 1.045 | 1.045 | 1.045 | 1.045 |
| Symmetry imposed | *C1* | *C1* | *C1* | *C1* |
| Initial particle images (no.) | 1,043,262 | 1,873,492 | 5,110,358 | 581,868 |
| Final particle images (no.) | 229,111 | 379,909 | 497,307 | 41,710 |
| Map resolution (Å) | 4.0 (3.9/4.4)* | 3.9 (3.7/4.0)* | 3.8 (3.6/3.9)* | 7.6 (6.8/6.8)* |
| FSC threshold | 0.143 | 0.143 | 0.143 | 0.143 |
| Map resolution range (Å) | 250.0-3.9 | 250.0-3.7 | 250.0-3.6 | 250.0-6.8 |
| | | | | |
| **Refinement** | | | | |
| Initial model used (PDB code) | 5MNW, 4GJ4, 3UVJ | 5MNW, 4GJ4, 3UVJ | 5MNW, 4GJ4, 3UVJ | |
| Model resolution (Å) | 4.0 | 3.8 | 3.8 | |
| FSC threshold | 0.5 | 0.5 | 0.5 | |
| Model resolution range (Å) | 250.0-4.0 | 250.0-3.8 | 250.0-3.8 | |
| Map sharpening $B$ factor (Å$^2$) | (-145/-219)* | (-167/-223)* | (-169/-210)* | |
| Model composition | | | | |
| Non-hydrogen atoms | 8,602 | 8,602 | 8,192 | |
| Protein residues | 1,118 | 1,118 | 1,075 | |
| Ligands | 1 | 1 | 4 | |
| $B$ factors (Å$^2$) | 172.24 | 160.29 | 133.90 | |
| Protein | 172.68 | 160.66 | 134.40 | |
| Ligand | 84.88 | 84.74 | 81.19 | |
| R.m.s. deviations | | | | |
| Bond lengths (Å) | 0.004 | 0.004 | 0.003 | |
| Bond angles (°) | 0.699 | 0.760 | 0.663 | |
| Validation | | | | |
| MolProbity score | 2.47 | 2.45 | 2.24 | |
| Clashscore | 10.91 | 10.27 | 10.08 | |
| Poor rotamers (%) | 4.73 | 4.62 | 3.54 | |
| Ramachandran plot | | | | |
| Favored (%) | 93.91 | 93.82 | 95.71 | |
| Allowed (%) | 5.73 | 5.82 | 4.09 | |
| Disallowed (%) | 0.36 | 0.36 | 0.29 | |

*The numbers outside the brackets are from the consensus refinement. Numbers inside brackets are from the multibody refinement (N-lobe/C-lobe).

# nature research

Corresponding author(s):   Lei Chen

Last updated by author(s):   2019/08/20

# Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

## Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

| n/a | Confirmed | |
|---|---|---|
| ☐ | ☒ | The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement |
| ☐ | ☒ | A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly |
| ☒ | ☐ | The statistical test(s) used AND whether they are one- or two-sided<br>*Only common tests should be described solely by name; describe more complex techniques in the Methods section.* |
| ☒ | ☐ | A description of all covariates tested |
| ☒ | ☐ | A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons |
| ☐ | ☒ | A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals) |
| ☒ | ☐ | For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted<br>*Give P values as exact values whenever suitable.* |
| ☒ | ☐ | For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings |
| ☒ | ☐ | For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes |
| ☒ | ☐ | Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated |

*Our web collection on statistics for biologists contains articles on many of the points above.*

## Software and code

Policy information about availability of computer code

| Data collection | SerialEM 3.6.11 |
|---|---|
| Data analysis | MotionCor2, GCTF, Gautomatch, RELION 3.0, cisTEM, cryoSPARC, PHENIX, Coot, UCSF Chimera, Pymol, GraphPad Prism 6, Microsoft Excel, caver |

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

## Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Cryo-EM maps and atomic coordinates of the heme-unliganded, heme-oxidized, NO-activated and β1 H105C mutant sGC have been deposited in the EMDB and PDB under the ID codes EMDB: EMD-9883, EMD-9884, EMD-9885, EMD-9886 and PDB: 6JT0, 6JT1, 6JT2, respectively.

# Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

# Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

| Sample size | No predetermination of sample size was performed. Sufficient cryo-EM data were collected to achieve adequate map resolutions for model building. The enzymatic activity assay experiments were performed with three biological replicates. The sample size was based on previous studies in the field and clearly indicated in the legends. |
|---|---|
| Data exclusions | Cryo-EM micrographs with ice or ethane contamination, empty carbon, and poor CTF fit (> 5 Å) were excluded manually. Particles belonging to bad classes were discarded and the data processing flowchart were summarized in Extended Data Fig. 1f, 3c. These criteria were pre-established and the procedure is a common practise in cryo-EM image analysis. No data was excluded in functional studies. |
| Replication | All attempts at replication were successful according to the detailed protocol described in the methods section. The numbers of replication were described in figure legends. |
| Randomization | For cryo-EM 3D refinement, all particles were randomly split into two groups. |
| Blinding | The investigators were blinded to group allocation during cryo-EM data collection and analysis. |

# Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

## Materials & experimental systems

| n/a | Involved in the study |
|---|---|
| ☒ | ☐ Antibodies |
| ☐ | ☒ Eukaryotic cell lines |
| ☒ | ☐ Palaeontology |
| ☒ | ☐ Animals and other organisms |
| ☒ | ☐ Human research participants |
| ☒ | ☐ Clinical data |

## Methods

| n/a | Involved in the study |
|---|---|
| ☒ | ☐ ChIP-seq |
| ☒ | ☐ Flow cytometry |
| ☒ | ☐ MRI-based neuroimaging |

# Eukaryotic cell lines

Policy information about cell lines

| Cell line source(s) | FreeStyle 293-F and Sf9 were from Thermo Fisher Scientific. |
|---|---|
| Authentication | None of the cell line used was authenticated. |
| Mycoplasma contamination | All cell lines were tested negative for mycoplasma contamination. |
| Commonly misidentified lines (See ICLAC register) | No commonly misidentified cell lines were used. |

# LETTER

# Stellar mergers as the origin of magnetic massive stars

Fabian R. N. Schneider[1,2,3,7]*, Sebastian T. Ohlmann[2,4,7]*, Philipp Podsiadlowski[3], Friedrich K. Röpke[2,5], Steven A. Balbus[3], Rüdiger Pakmor[6] & Volker Springel[6]

**About ten per cent of 'massive' stars (those of more than 1.5 solar masses) have strong, large-scale surface magnetic fields[1–3]. It has been suggested that merging of main-sequence and pre-main-sequence stars could produce such strong fields[4,5], and the predicted fraction of merged massive stars is also about ten per cent[6,7]. The merger hypothesis is further supported by a lack of magnetic stars in close binaries[8,9], which is as expected if mergers produce magnetic stars. Here we report three-dimensional magnetohydrodynamical simulations of the coalescence of two massive stars and follow the evolution of the merged product. Strong magnetic fields are produced in the simulations, and the merged star rejuvenates such that it appears younger and bluer than other coeval stars. This can explain the properties of the magnetic 'blue straggler' star $\tau$ Sco in the Upper Scorpius association that has an observationally inferred, apparent age of less than five million years, which is less than half the age of its birth association[10]. Such massive blue straggler stars seem likely to be progenitors of magnetars, perhaps giving rise to some of the enigmatic fast radio bursts observed[11], and their supernovae may be affected by their strong magnetic fields[12].**

We conduct three-dimensional (3D) ideal magnetohydrodynamical (MHD) simulations of the merger of a 9-Myr-old binary consisting of a $9M_\odot$ and an $8M_\odot$ (where $M_\odot$ is the solar mass) core-hydrogen burning star with the moving-mesh code AREPO[13], which is ideally suited for such simulations (see Methods). The binary configuration and evolutionary stage are chosen such that the resulting merger product is expected to have a total mass similar to that of $\tau$ Sco (about $17M_\odot$; ref. [10]) and that the binary could have formed at the same time as did other stars in the Upper Scorpius association.

Snapshots of the density $\rho$, a passive scalar indicating material from the primary star, and of the absolute magnetic-field strength $B$ of the 3D MHD simulation are shown in Fig. 1. Upon contact of the binary, a dynamical phase of mass transfer with rates as high as $17M_\odot$ yr$^{-1}$ sets in from the more massive (initially $9M_\odot$) primary star onto the less massive (initially $8M_\odot$) secondary star. Mass is lost through the outer Lagrangian points, draining angular momentum and thereby accelerating the coalescence. The accretion stream shears on the surface of the secondary star and it is in this accretion stream, of size 0.8 solar radii ($R_\odot$), that the magnetic field is amplified on an e-folding timescale of about 0.2–1 d (Fig. 1d, g). The maximum magnetic-field strength saturates at about $10^8$ G, which corresponds to an amplification of the magnetic energy by a factor of $10^{18}$ (Extended Data Fig. 2). At saturation, the magnetic energy is comparable to the turbulent energy (about 5%–30%), which is the source of the magnetic-field amplification (Extended Data Fig. 3). In the final merger, the amplified field is advected throughout the merger product and is therefore also present in the core of the merger remnant. When the primary star is disrupted around the secondary and the cores of the two stars merge, vortices form at the interface of the two former cores (Fig. 1e) that further contribute to the magnetic-field amplification (see also Supplementary

Video 1). The maximum ratio of magnetic to gas pressure reaches 30% in localized regions but is less than 1% in the phase leading up to the merger.

The local conditions in the differentially rotating accretion stream (rotational frequency of $\Omega \approx 10$ d$^{-1}$, Alfvén velocity of about 1 km s$^{-1}$ and rotational shear of $q = -\mathrm{d}\ln\Omega/\mathrm{d}\ln r \approx 0.4$) indicate that the magneto-rotational instability[14] is the key agent providing the turbulence needed to exponentially amplify the magnetic fields. In the shearing layer, the fastest-growing mode of the magneto-rotational instability has a characteristic size of $0.1R_\odot$ and growth timescale of 0.5 d (ref. [15]), in agreement with the size of the accretion stream and the observed growth timescale of the magnetic fields in our simulation.

Because of the large amount of angular momentum, a torus of $3M_\odot$ forms that surrounds the central, spherically symmetric $14M_\odot$ core of the merger product (Fig. 1c, l). The central merger remnant is in solid-body rotation while the centrifugally supported torus rotates at sub-Keplerian velocities. The innermost core of the merger remnant consists of material from the former secondary star while the torus is dominated by core material from the former primary star (Fig. 1f).

We continue the 3D MHD simulation for 10 d after the actual merger, that is, about 5 d after the merger remnant has settled into its final core-torus structure. This corresponds to roughly 5 Alfvén crossing timescales through the $14M_\odot$ core and we do not observe large changes in the magnetic field structure and strength. The ratio of toroidal to total magnetic field energy is 80%–85%, which is in a regime where magnetic-field configurations are thought to be stable in stellar interiors[16]. Because of the high conductivity of stellar plasmas, Ohmic decay of the field occurs only on a timescale similar to or even longer than the stellar lifetime (see Methods). We therefore expect the magnetic field to be long-lived.

Most of the torus is expected to be accreted rapidly onto the central merger remnant and to form an extended envelope (see Methods). For the long-term evolution of the merger remnant we therefore assume that the innermost $16.9M_\odot$ end up in the merger product and that less than $0.1M_\odot$ remain in a disk (Extended Data Fig. 1). Under these assumptions, we follow the further evolution of the merger product in the one-dimensional (1D) stellar evolution code Mesa[17]. As suggested by the 3D MHD simulations, we assume the formed remnant to rotate rigidly at a rate close to break-up. The magnetic flux at the end of our 3D simulation at a mass coordinate of $16.9M_\odot$ is $3.5 \times 10^{23}$ G cm$^2$, such that the surface magnetic-field strength of the merger remnant on the main sequence would be 9 kG for a radius of about $5R_\odot$ (assuming magnetic flux conservation). This is well within observed surface field strengths of magnetic stars[1–3]. Because it is impossible to follow the evolution of an inherently 3D magnetic field in a 1D stellar evolution code, we assume that the radial magnetic-field strength in our 1D model follows that of a magnetic dipole. It contributes to internal angular-momentum transport and additional angular-momentum loss from the

[1]Zentrum für Astronomie der Universität Heidelberg, Astronomisches Rechen-Institut, Heidelberg, Germany. [2]Heidelberger Institut für Theoretische Studien, Heidelberg, Germany. [3]Department of Physics, University of Oxford, Oxford, UK. [4]Max Planck Computing and Data Facility, Garching, Germany. [5]Zentrum für Astronomie der Universität Heidelberg, Institut für Theoretische Astrophysik, Heidelberg, Germany. [6]Max-Planck-Institut für Astrophysik, Garching, Germany. [7]These authors contributed equally: Fabian R. N. Schneider, Sebastian T. Ohlmann. *e-mail: fabian.schneider@uni-heidelberg.de; sebastian.ohlmann@mpcdf.mpg.de
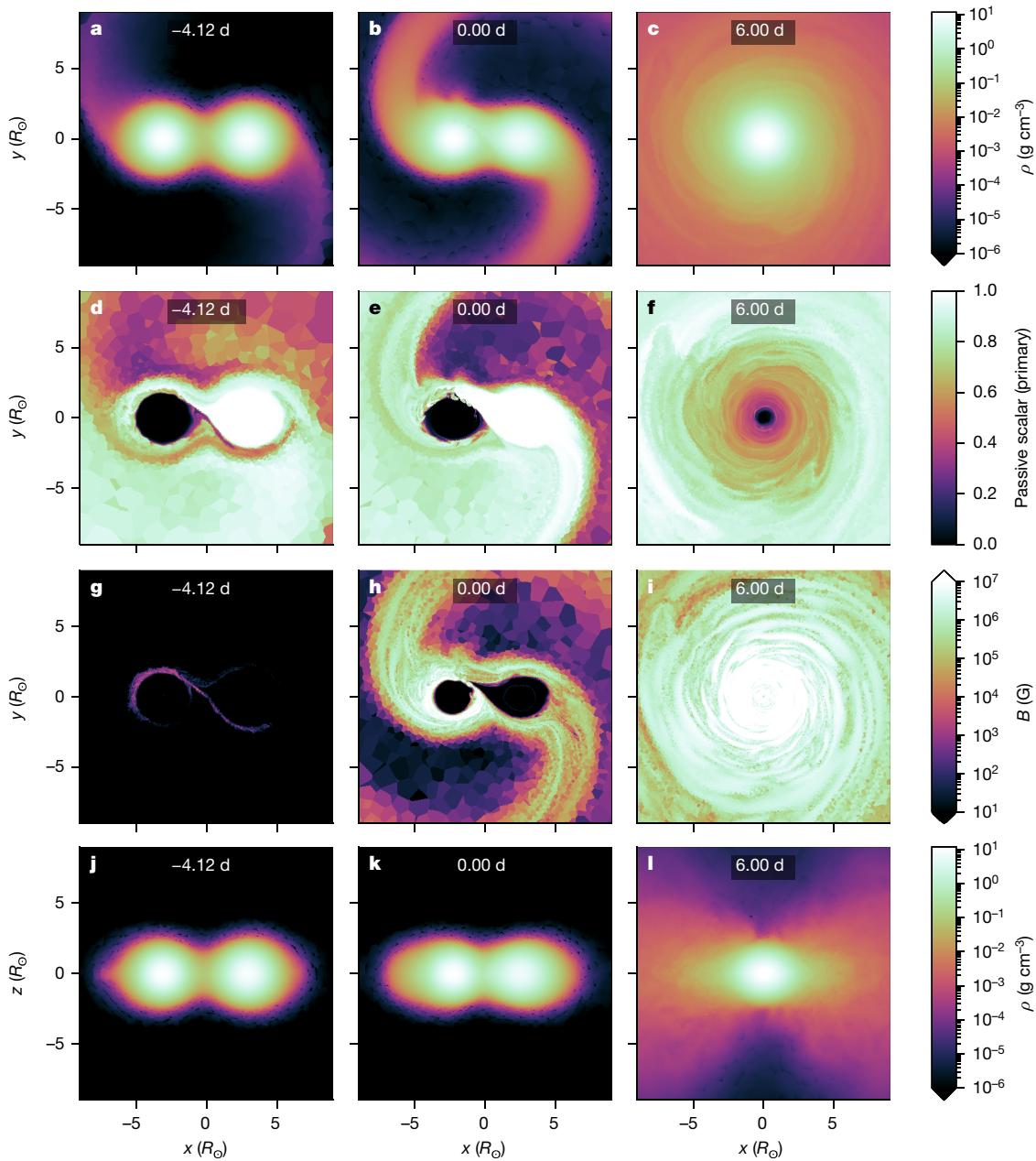
**Fig. 1 | Dynamical evolution of the merger of two main-sequence stars.**
Panels **a**–**c** show density snapshots in the orbital plane; panels **j**–**l** are
edge-on views of the density. The passive scalar (white colour; panels **d**–**f**)
indicates material from the $9M_\odot$ primary and thus visualizes the mixing
of the two progenitor stars during the merger. The passive scalar and the
magnetic-field strengths (panels **g**–**i**) are shown in the orbital plane. The
times given in each panel are relative to the time when the cores of the two
stars coalesce (panels **b**, **e**, **h**, and **k**).

surface through a magnetized stellar wind (magnetic braking) but has
otherwise no influence on the structure and evolution (see Methods).

Because of the coalescence, the stellar interior is heated and the star is
out of thermal equilibrium. A thermal relaxation phase sets in, during
which the star reaches a maximum radius of about $200R_\odot$ and a lumi-
nosity of $2.5 \times 10^5 L_\odot$, (where $L_\odot$ is the solar luminosity) before it con-
tracts back to the main sequence, after which it continues its evolution
in a manner similar to that of a genuine single star of initially $16.9M_\odot$
(Fig. 2). During the thermal expansion, the star reaches critical rotation,
leading to additional mass (less than $0.01M_\odot$) and angular-momentum
loss (roughly 7% of the star's total angular momentum). In the subse-
quent contraction phase, the surface spins down from critical rotation
to about 50 km s$^{-1}$ (about 10% of critical rotation). This is not driven by
angular-momentum loss but by an internal restructuring of the star (see
also Methods). The spin of the merger product on the main sequence is
thus set by the angular momentum that remains in the merger product

after the viscous accretion of the torus and the corresponding outward
angular-momentum transport.

Once back in thermal equilibrium, the merger product is a slow
rotator with effective temperature, luminosity and surface gravity in
agreement with $\tau$ Sco (Table 1 and Fig. 2). This outcome is independ-
ent of the magnetic field chosen in the 1D models. Consequently, our
merger product also looks like a rejuvenated blue straggler compared to
other, apparently older stars in the Upper Scorpius association, mainly
because of the shorter lifetime associated with the now more-massive
star. $\tau$ Sco is enriched in nitrogen on the surface but this is currently
not reproduced by our model. However, on average, the envelope of our
merger model is nitrogen-rich because it is made out of core material
of the former primary star (Fig. 1f). These enriched layers could easily
be exposed by further mass loss or could be mixed to the stellar surface.
For example, we have not considered mixing induced by the magnetic
fields or during the viscous evolution of the torus. In conclusion, our
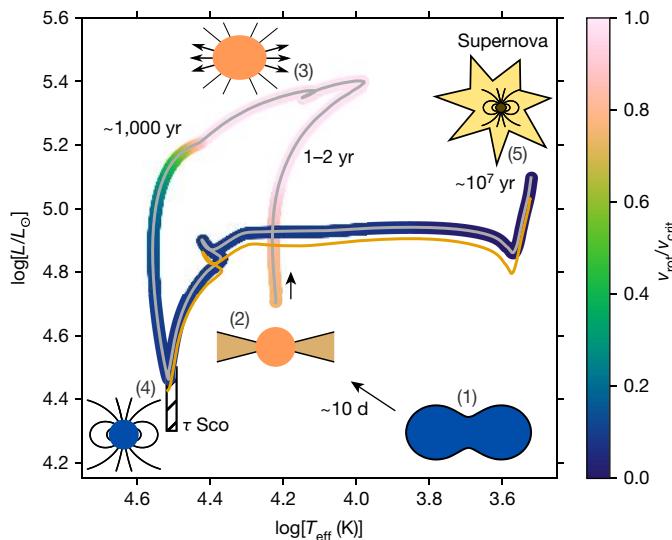
**Fig. 2 | Long-term evolution of the merger product in the Hertzsprung–Russell diagram.** After most of the torus is accreted, the merger remnant rotates rapidly and a thermal relaxation phase sets in, during which the star first expands before it contracts back to the main sequence (grey line). The direction of evolution is indicated by the black arrow at the beginning of this phase. The colour-coding shows the surface rotational velocity, $v_{rot}$, in terms of the critical Keplerian velocity, $v_{crit}$. Once on the main sequence, the merger product continues its evolution similar to that of a genuine single star of the same mass of $16.9M_\odot$ (orange line). The black hatched rectangle indicates observations of $\tau$ Sco ($31{,}000$ K $\leq T_{eff} \leq 33{,}000$ K, $4.3 \leq \log L/L_\odot \leq 4.5$; see Table 1). The small cartoons are artist's impressions of key evolutionary phases: (1) the contact phase before the actual merger, (2) the merger product with its torus, (3) during the thermal relaxation as a critically rotating star shedding mass, (4) as a main-sequence star with a strong surface magnetic field and (5) after the terminal supernova explosion that may form a magnetar.

merger scenario is able to explain the magnetic nature, atmospheric parameters, slow rotation and blue-straggler status of $\tau$ Sco.

Strong amplification of magnetic fields is also observed in the coalescence of white dwarfs[18], the merger of neutron stars[19], and the common-envelope phase of a star spiralling into the envelope of a giant companion[20]. Thus, mergers of stars in general seem to provide the right conditions to produce strong magnetic fields. The coalescence of other main-sequence stars (having, for example, lower masses) with stars in different evolutionary phases (such as pre-main-sequence stars) is also expected to generate strong magnetic fields. Merging is therefore also a promising mechanism with which to explain the origin of magnetic fields in Ap stars and their suggested remnants, white dwarfs with surface field strengths in excess of $10^6$ G (ref. [5]).

The magnetic flux in the innermost $1.5M_\odot$ of our merger model at the end of the MHD simulation is about $4 \times 10^{28}$ G cm$^2$. If all of the magnetic flux is conserved until core collapse of the merger product occurs, a resulting neutron star of radius 10 km would have a surface magnetic-field strength of about $10^{16}$ G. Such strong fields are thought to affect the explosions of core-collapse supernovae[12] and appear to be able to explain the strong magnetic fields inferred for magnetars

$(10^{13}$–$10^{15}$ G$)^{21}$, which may give rise to some of the enigmatic fast radio bursts[11]. The birthrate of magnetars in our Galaxy of about 0.3 per century[22] and the rate of occurrence of Galactic core-collapse supernovae of about 2 per century[23] suggest that 15% of all Galactic core-collapse supernovae have produced a magnetar, which is consistent with the 10% incidence of magnetic massive stars[24]. Super-luminous supernovae and long-duration gamma-ray bursts have been suggested to be powered by rapidly rotating and highly magnetized cores[25,26]. Because of its slow rotation, our merger model is not expected to result in such events, in line with the low rate of these transients (less than 0.1% of core-collapse supernovae)[24]. However, rarer merger cases such as the coalescence of stellar cores in a common-envelope event[27] could form rapidly rotating and highly magnetized stellar cores that may then power long-duration gamma-ray bursts and super-luminous supernovae. Taken together, this enables our merger model to explain the strong magnetic fields observed in a subset of massive stars and potentially also the origin of magnetars.

## Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-019-1621-5.

1. Donati, J.-F. & Landstreet, J. D. Magnetic fields of nondegenerate stars. *Annu. Rev. Astron. Astrophys.* **47**, 333–370 (2009).
2. Fossati, L. et al. B fields in OB stars (BOB): low-resolution FORS2 spectropolarimetry of the first sample of 50 massive stars. *Astron. Astrophys.* **582**, A45 (2015).
3. Grunhut, J. H. et al. The MiMeS survey of magnetism in massive stars: magnetic analysis of the O-type stars. *Mon. Not. R. Astron. Soc.* **465**, 2432–2470 (2017).
4. Ferrario, L., Pringle, J. E., Tout, C. A. & Wickramasinghe, D. T. The origin of magnetism on the upper main sequence. *Mon. Not. R. Astron. Soc.* **400**, L71–L74 (2009).
5. Wickramasinghe, D. T., Tout, C. A. & Ferrario, L. The most magnetic stars. *Mon. Not. R. Astron. Soc.* **437**, 675–681 (2014).
6. Podsiadlowski, P., Joss, P. C. & Hsu, J. J. L. Presupernova evolution in massive interacting binaries. *Astrophys. J.* **391**, 246–264 (1992).
7. de Mink, S. E., Sana, H., Langer, N., Izzard, R. G. & Schneider, F. R. N. The incidence of stellar mergers and mass gainers among massive stars. *Astrophys. J.* **782**, 7 (2014).
8. Carrier, F., North, P., Udry, S. & Babel, J. Multiplicity among chemically peculiar stars. II. Cool magnetic Ap stars. *Astron. Astrophys.* **394**, 151–169 (2002).
9. Alecian, E. et al. The BinaMIcS project: understanding the origin of magnetic fields in massive stars through close binary systems. In *IAU Symposium* Vol. 307, 330–335 (IAU, 2015).
10. Schneider, F. R. N., Podsiadlowski, P., Langer, N., Castro, N. & Fossati, L. Rejuvenation of stellar mergers and the origin of magnetic fields in massive stars. *Mon. Not. R. Astron. Soc.* **457**, 2355–2365 (2016).
11. Metzger, B. D., Berger, E. & Margalit, B. Millisecond magnetar birth connects FRB 121102 to superluminous supernovae and long-duration gamma-ray bursts. *Astrophys. J.* **841**, 14 (2017).
12. Obergaulinger, M., Janka, H. T. & Aloy, M. A. Magnetic field amplification and magnetically supported explosions of collapsing, non-rotating stellar cores. *Mon. Not. R. Astron. Soc.* **445**, 3169–3199 (2014).
13. Springel, V. E pur si muove: Galilean-invariant cosmological hydrodynamical simulations on a moving mesh. *Mon. Not. R. Astron. Soc.* **401**, 791–851 (2010).
14. Balbus, S. A. & Hawley, J. F. A powerful local shear instability in weakly magnetized disks. I—Linear analysis. II—Nonlinear evolution. *Astrophys. J.* **376**, 214–233 (1991).
15. Rembiasz, T., Obergaulinger, M., Cerdá-Durán, P., Müller, E. & Aloy, M. A. Termination of the magnetorotational instability via parasitic instabilities in core-collapse supernovae. *Mon. Not. R. Astron. Soc.* **456**, 3782–3802 (2016).
16. Braithwaite, J. & Nordlund, Å. Stable magnetic fields in stellar interiors. *Astron. Astrophys.* **450**, 1077–1095 (2006).
17. Paxton, B. et al. Modules for Experiments in Stellar Astrophysics (MESA). *Astrophys. J. Suppl. Ser.* **192**, 3 (2011).
18. Zhu, C., Pakmor, R., van Kerkwijk, M. H. & Chang, P. Magnetized moving mesh merger of a carbon-oxygen white dwarf binary. *Astrophys. J.* **806**, L1 (2015).
19. Price, D. J. & Rosswog, S. Producing ultrastrong magnetic fields in neutron star mergers. *Science* **312**, 719–722 (2006).
20. Ohlmann, S. T., Röpke, F. K., Pakmor, R., Springel, V. & Müller, E. Magnetic field amplification during the common envelope phase. *Mon. Not. R. Astron. Soc.* **462**, L121–L125 (2016).
21. Olausen, S. A. & Kaspi, V. M. The McGill Magnetar Catalog. *Astrophys. J. Suppl. Ser.* **212**, 6 (2014).

## Table 1 | Comparison of our merger model with $\tau$ Sco

| | $T_{eff}$ (K) | $\log[L/L_\odot]$ | $\log[g\ (cm\ s^{-2})]$ |
|---|---|---|---|
| Merger model | $\sim 32{,}500$ | $\sim 4.50$ | $\sim 4.17$ |
| Ref. [28] | $31{,}900^{+500}_{-800}$ | $4.39 \pm 0.09$ | $4.15^{+0.09}_{-0.14}$ |
| Ref. [29] | $32{,}000 \pm 1{,}000$ | $4.47 \pm 0.13$ | $4.00 \pm 0.10$ |
| Ref. [30] | $32{,}000 \pm 300$ | $4.33 \pm 0.05$ | $4.33 \pm 0.06$ |

The effective temperature ($T_{eff}$), luminosity ($\log L/L_\odot$) and surface gravity ($\log g$) of the merger product after thermal relaxation are compared to three sets of observations of $\tau$ Sco. The provided uncertainties are 68.3% confidence limits.

22. Keane, E. F. & Kramer, M. On the birthrates of Galactic neutron stars. *Mon. Not. R. Astron. Soc.* **391**, 2009–2016 (2008).
23. Diehl, R. et al. Radioactive ²⁶Al from massive stars in the Galaxy. *Nature* **439**, 45–47 (2006).
24. Moriya, T. J., Sorokina, E. I. & Chevalier, R. A. Superluminous supernovae. *Space Sci. Rev.* **214**, 59 (2018).
25. Bucciantini, N. et al. Magnetized relativistic jets and long-duration GRBs from magnetar spin-down during core-collapse supernovae. *Mon. Not. R. Astron. Soc.* **396**, 2038–2050 (2009).
26. Kasen, D. & Bildsten, L. Supernova light curves powered by young magnetars. *Astrophys. J.* **717**, 245–249 (2010).
27. Tout, C. A., Wickramasinghe, D. T., Lau, H. H.-B., Pringle, J. E. & Ferrario, L. A common envelope binary star origin of long gamma-ray bursts. *Mon. Not. R. Astron. Soc.* **410**, 2458–2462 (2011).
28. Mokiem, M. R. et al. Spectral analysis of early-type stars using a genetic algorithm based fitting method. *Astron. Astrophys.* **441**, 711–733 (2005).
29. Simón-Díaz, S., Herrero, A., Esteban, C. & Najarro, F. Detailed spectroscopic analysis of the Trapezium cluster stars inside the Orion nebula. Rotational velocities, stellar parameters, and oxygen abundances. *Astron. Astrophys.* **448**, 351–366 (2006).
30. Nieva, M.-F. & Przybilla, N. Fundamental properties of nearby single early B-type stars. *Astron. Astrophys.* **566**, A7 (2014).

## METHODS

**3D MHD merger simulations.** *AREPO's MHD solver.* The AREPO code uses a second-order finite-volume method to solve the ideal MHD equations on an unstructured grid[13,31,32]. The grid is generated in each timestep from a set of mesh-generating points that move along with the flow, thus ensuring a nearly Lagrangian behaviour while regularizing the mesh by adding an additional term. The fluxes over the cell boundaries are computed using the HLLD solver and the divergence of the magnetic field is effectively controlled by employing the Powell scheme[33], as shown in refs 31, 32. We use ideal MHD here because the resistivity is very small in the highly conducting plasma of stellar interiors (see Ohmic dissipation below).

*Initialization of the binary progenitor.* The binary progenitors have an initial helium mass fraction of $Y = 0.2703$ and solar metallicity $Z = 0.0142$ (ref. [34]). The stellar structures are imported from 1D Mesa[17,35–37] models (version 9793) that employ exponential convective-core overshooting with a parameter of $f_{ov} = 0.019$, which effectively corresponds to a step convective-core overshooting of about 0.16 pressure scale heights.

Mapping the 1D stellar structures into a 3D hydrodynamics code leads to discretization errors in the hydrostatic equilibrium; thus, a relaxation method[38] is employed to create stable stellar models. The 1D stellar models from MESA are mapped onto an unstructured grid consisting of HEALPix distributions on spherical shells[39]. In the ensuing AREPO simulations, spurious velocities are damped away, resulting in stable stellar models according to the criteria outlined in ref. [38]. The initial seed magnetic field was set up in a dipole configuration with a polar surface field strength of 1 µG.

The relaxed single-star models are subsequently used to set up the binary star merger. It is computationally not feasible to simulate the merger beginning from Roche-lobe overflow until the actual merger occurs. We therefore speed up the merging process by artificially decelerating each cell for a certain time (about 1.5 orbits) and starting the calculation from then. The duration of this deceleration phase influences the outcome of the merger only marginally (see resolution study below).

*Resolution study and initial conditions.* We ran simulations for different resolutions and initial binary setups to ensure that the amplification of the magnetic field is robust against variations of these parameters. The evolution of the total magnetic field energy over time is shown in Extended Data Fig. 2. The standard run shown in the main text is Model 1. The evolution of the magnetic energy is slightly different for the various configurations but the overall behaviour and the final energy are very similar. Model 2 tests a lower resolution for otherwise identical initial conditions. Model 3 was started at an earlier time with a larger initial separation. The resolution was set up with roughly $4 \times 10^6$ cells for Model 1 and about $4 \times 10^5$ cells for the other models.

*Magnetic-field saturation.* The magnetic-field amplification switches off if the necessary physical conditions of the amplification process are no longer met. For the magneto-rotational instability[14], this could be the case if the magnetic field becomes so strong that the fastest-growing mode exceeds the spatial region of interest (for example, it becomes larger than the star), if the amplification timescale becomes excessively long or if there is no longer differential rotation. Such a situation may go along with an equipartition of the magnetic energy with the energy source (for example, differential or turbulent energy) that drives the magnetic-field amplification.

In our models, the initially fast, exponential magnetic-field amplification is consistent with being driven by the magneto-rotational instability. After the merger, the central star is in solid-body rotation such that the magneto-rotational instability cannot operate within the central star any more. In the torus, however, the magneto-rotational instability is still active and the magnetic-field strength is indeed found to increase until the end of the simulation. The fastest-growing magneto-rotational instability mode always fits into the central merger remnant and the magneto-rotational instability amplification timescale stays short compared to the runtime of our simulation.

Using the kinetic energy in the radial and $z$ directions as a proxy for the turbulent energy, which is generally thought to power the magnetic-field amplification, we find that the magnetic energy reaches a level of about 5%–30% of the kinetic energy (Extended Data Fig. 3). This supports the idea that the magnetic-field amplification ceases when approaching energy equipartition.

*Ohmic dissipation of magnetic fields.* Stable magnetic fields can diffuse out of the stellar interior by Ohmic resistivity and thereby dissipate[16]. However, because the hot stellar interior is highly conducting, the resistivity is low and the dissipation of magnetic fields is slow. Indeed, for Spitzer's resistivity[40]:

$$\eta = 7 \times 10^{11} \ln \Lambda (T)^{-3/2} \mathrm{cm}^2 \mathrm{s}^{-1} \tag{1}$$

with $T$ the temperature in Kelvin and $\ln \Lambda$ the Coulomb logarithm, which is of the order of 10 for stellar interiors. The diffusion timescale of magnetic fields

is $\tau_{diff} = R^2/\eta \approx 10^8$–$10^{11}$ yr for temperatures of $10^5$–$10^7$ K and a length scale of $R = 1 R_\odot$. These estimates depend on the still-uncertain resistivity in stellar interiors but it appears that Ohmic dissipation of the amplified magnetic fields does not play a part in this merger because the lifetime of the merger product is instead of the order of $10^7$ yr. It might, however, be relevant for some evolved stars[16].

**1D long-term evolution of the merger product.** The amplified magnetic fields are too weak to affect the stellar structure directly. However, they can contribute to the angular-momentum transport through the stellar interior and may lead to additional angular-momentum loss from the stellar surface (magnetic braking). Below, we describe the assumed magnetic field structure in our 1D stellar evolution models, and our implementation of the interior angular-momentum transport through the magnetic fields and magnetic braking. We then explain how our 1D models are set up, on the basis of the outcome of the 3D MHD simulations and provide more details on the spin-down of the merger product. As before, we use the Mesa stellar evolution code in version 9793[17,35–37].

*Assumed large-scale magnetic field in 1D computations.* It is not possible to follow the evolution of a 3D magnetic field in a 1D stellar evolution code. Moreover, the final configuration of the magnetic field after the accretion of the torus is uncertain at present. Hence, we assume that the radial magnetic-field strength in our 1D model follows that of a magnetic dipole, $B(r) = \mu_B r^{-3}$, (where $r$ is the radius of the product of the merged stars) with dipole moment $\mu_B = 2 \times 10^{37}$ G cm³. This assumption is conservative in the sense that it results in a surface magnetic field of the merger product on the main sequence of a few hundred Gauss, which is lower than that expected from magnetic flux freezing of our 3D model but reminiscent of that of $\tau$ Sco[41]. Using larger or smaller magnetic-dipole moments does not affect our conclusions. The dipole field diverges for $r \to 0$ and we therefore cap its field strength at $10^9$ G.

We further assume that the magnetic field is expelled from convective regions if the convective energy density $u_{conv}$ is larger than the magnetic energy density $u_B$, that is, if:

$$u_{conv} = \frac{1}{2} \rho v_{conv}^2 > u_B = \frac{B^2}{8\pi} \tag{2}$$

Here, $\rho$ is the gas density and $v_{conv}$ is the velocity of convective eddies as predicted by mixing-length theory. This treatment of the static magnetic field means that it contributes only to the angular-momentum transport in radiative regions.

*Angular-momentum transport in the stellar interior through a large-scale magnetic field.* We treat the transport of angular momentum through the stellar interior as a diffusive process. Magnetic fields cause Maxwell stresses and can thus transport angular momentum. To obtain the effective diffusion coefficient of this process (which we call effective viscosity, $\nu_{eff}$), we consider differentially rotating, spherical shells and assume that the stresses due to magnetic fields are effectively similar to the classical Newtonian dynamic shear, $S$:

$$S = \frac{dF}{dA} = \nu_{eff} \rho \frac{\partial v}{\partial r} \tag{3}$$

where $dF$ is the force exerted by the shear on an area $dA$ and $\partial v/\partial r$ is the radial gradient of the velocity $v$. In spherical coordinates $(r, \varphi, \theta)$, the torque $d\tau$ on a surface element $dA = r^2 \sin\theta d\varphi d\theta$ due to a shear force $dF$ is:

$$d\tau = r\sin\theta dF = r\sin\theta \nu_{eff} \rho \frac{\partial v}{\partial r} dA \tag{4}$$

Introducing the angular velocity $\Omega$ ($v = r\sin\theta\Omega$), we have $\partial v/\partial \Omega = r\sin\theta$ and thus $\partial v/\partial r = r\sin\theta \partial \Omega/\partial r$. Integrating equation (4) over $\varphi$ and $\theta$, we obtain the overall torque on a shell at radius $r$ as:

$$\tau = \frac{8\pi}{3} \nu_{eff} r^4 \rho \frac{\partial \Omega}{\partial r} \tag{5}$$

From a physical point of view, the shear exerted by magnetic fields reduces differential rotation and attempts to establish solid-body rotation ($\partial \Omega/\partial r = 0$). The amount of angular momentum $\Delta J$ that needs to be transported to achieve solid-body rotation in neighbouring, differentially rotating shells is $\Delta J = I \Delta \Omega$, where $I$ is the moment of inertia. The angular-momentum transport across a shell of thickness $\Delta r$ occurs with an Alfvén velocity $v_A = B/\sqrt{4\pi\rho}$, that is, on an Alfvén timescale of $\tau_A = \Delta r/v_A$, such that:

$$\frac{dJ}{dt} \approx \frac{\Delta J}{\tau_A} = \frac{I(\partial \Omega/\partial r)\Delta r}{\tau_A} = I\left(\frac{\partial \Omega}{\partial r}\right)v_A \tag{6}$$

Equating equation (6) and equation (5), we find the desired effective viscosity for angular-momentum transport in shells rotating differentially because of a large-scale magnetic field:

$$\nu_{\text{eff}} = \frac{3I}{8\pi r^4 \rho} \nu_{\text{A}} \tag{7}$$

The moment of inertia of a single shell in a stellar evolution model depends on the spatial discretization. To make the effective diffusion coefficient independent of resolution of the spatial discretization of the stellar model, we define 'shells' to have a thickness of 20% of the local pressure scale height $H_{\text{P}}$. We modulate the effective viscosity with a factor $f_{\text{A}}$ that is thought to adjust the timescale over which solid-body rotation is achieved in neighbouring shells. We set $f_{\text{A}} = 0.5$ in our calculations and note that small variations in $f_{\text{A}}$ hardly change our results.

In the above analysis, we have not made explicit assumptions about the magnetic-field geometry, but it will of course matter in reality. For example, if there is no radial magnetic-field component, the Maxwell stress is zero, such that there is no angular-momentum transport in the radial direction through the magnetic field. In our approach, the field geometry enters indirectly through the Alfvén velocity, which depends on the absolute magnetic-field strength, which itself is a function of radius $r$.

*Magnetic braking.* Stellar winds can couple to large-scale magnetic fields and thereby enhance the loss of angular momentum, a process called magnetic braking. The torque on the stellar surface from magnetic braking is:

$$\frac{dJ_{\text{mb}}}{dt} = \frac{2}{3} \dot{M} \Omega_* R_{\text{A}}^2 \tag{8}$$

where $\dot{M}$ is the stellar wind mass loss rate, $\Omega_*$ is the stellar surface angular velocity and the factor 2/3 accounts for the moment of inertia of thin spherical shells.

In MHD simulations of magnetic braking of hot, massive stars[42], the Alfvén radius in equation (8) is found to be:

$$\frac{R_{\text{A}}}{R_*} \approx 0.29 + (\eta_* + 0.25)^{1/4} \tag{9}$$

with $R_*$ the stellar radius and $\eta_*$ the wind magnetic confinement parameter:

$$\eta_* = \frac{B_{\text{eq}}^2 R_*^2}{\dot{M} v_\infty} \tag{10}$$

where $B_{\text{eq}}$ is the equatorial, surface magnetic-field strength and $v_\infty$ is the terminal wind velocity[42]. For the terminal wind velocity, we use observational results for O to F stars[43].

From a technical point of view, stellar winds in stellar evolution codes take away the specific angular momentum of their former Lagrangian mass shells. In our models, the additional angular momentum lost through magnetic braking is then taken away from a thin surface layer after the mass shells that are lost in the wind have been removed.

*Import of the merger remnant into a 1D stellar evolution code.* Immediately after the merger, the evolution is driven mainly by that of the torus and its interplay with the central star. Two timescales are most relevant: the accretion and cooling timescales, $\tau_{\text{acc}}$ and $\tau_{\text{cool}}$, respectively. The accretion timescale sets the time over which the torus is accreted by the central remnant whereas the cooling timescale describes the time over which the torus loses the heat produced by the accretion.

Matter in the rotationally supported torus can only be accreted onto the central star if its angular momentum is transported outwards; hence, the accretion timescale is given by the angular-momentum transport timescale. We assume that the matter and angular momentum flow in the torus can be described by an $\alpha$-disk model with an effective viscosity $\alpha$ that, for example, might be provided by the magnetic fields or the magneto-rotational instability[14]. Here and throughout, the term 'viscous' is used in the phenomenological sense of an effective viscosity which acts on large scales owing to the presence of an enhanced turbulent transport. It should not be confused with the true microscopic particle viscosity, which is negligible for the problems of interest here. Using the mass accretion rate for such an $\alpha$-disk model:

$$\dot{M}_{\text{acc}} \approx 3\alpha \left(\frac{h}{r}\right)^2 \Omega M_{\text{disk}} \tag{11}$$

The accretion timescale of the torus is then:

$$\tau_{\text{acc}} = \frac{M_{\text{disk}}}{\dot{M}_{\text{acc}}} = \frac{1}{3} \frac{r^2}{\alpha h^2 \Omega} = 0.02 \, \text{yr} \left(\frac{10^{-2}}{\alpha}\right) \left(\frac{r/h}{2}\right)^2 \left(\frac{h^{-1}}{\Omega}\right) \tag{12}$$

Here, $h/r$ is the ratio of disk height and radius, $M_{\text{disk}}$ is the mass in the disk and $\Omega$ is the angular velocity of the disk, which generally depends on radius. In our case, the accretion is equivalent to the viscous timescale $\tau_{\text{visc}}$.

Mass accretion leads to (turbulent) heating through the release of gravitational potential energy, $E_{\text{grav}}$. On the one hand, if this energy can be lost efficiently from the system via fast cooling ($\tau_{\text{cool}} \ll \tau_{\text{acc}}$), the torus becomes thinner or at least keeps its shape. On the other hand, if the cooling is inefficient ($\tau_{\text{cool}} \gg \tau_{\text{acc}}$), the torus becomes thicker and evolves into a thermally supported extended envelope. Assuming that the star-torus structure radiates at a fraction $f_{\text{Edd}}$ of its Eddington luminosity, $L_{\text{Edd}}$, and that photon cooling is the dominant cooling process, the cooling timescale can be approximated as:

$$\tau_{\text{cool}} = \frac{E_{\text{grav}}}{f_{\text{Edd}} L_{\text{Edd}}} = \frac{1}{f_{\text{Edd}}} \frac{GM_{\text{core}} M_{\text{disk}}/R_{\text{core}}}{4\pi G(M_{\text{core}} + M_{\text{disk}})c/\kappa}$$
$$\approx 0.8 \times 10^3 \text{yr} \left(\frac{1}{f_{\text{Edd}}}\right) \left(\frac{1+X}{1.7}\right) \left(\frac{M_{\text{core}} M_{\text{disk}}/M_\odot}{M_{\text{core}} + M_{\text{disk}}}\right) \left(\frac{R_\odot}{R_{\text{core}}}\right) \tag{13}$$

where $M_{\text{core}}$ and $R_{\text{core}}$ are the mass and radius of the central star, and $\kappa$ is the opacity. In the last step, we assumed that the opacity is dominated by electron scattering, that is, $\kappa = 0.2(1 + X) \text{ cm}^2 \text{ g}^{-1}$ with the hydrogen mass fraction $X$. Even for $f_{\text{Edd}} = 1$, the cooling time in our case is of the order of 500–700 yr ($M_{\text{core}} = 14M_\odot$, $M_{\text{disk}} = 3M_\odot$, $3 \leq R_{\text{core}}/R_\odot \leq 4$) and thus much longer than the accretion timescale in equation (12). The expectation therefore is that the torus is rapidly accreted onto the central star and evolves into a thermally supported, extended envelope before its thermal relaxation and cooling process sets in.

These arguments are analogous to previous work on the merger remnant of two white dwarfs[44]. More detailed simulations of the viscous evolution of this double white-dwarf merger remnant[45] support the (analytic) expectations[44] and indeed show a rapid transformation of the torus into a thermally supported envelope. Given the similarity of the physical situation and the ratio of accretion and cooling timescales in our case, it seems reasonable that large fractions of our torus will also evolve into a thermally supported envelope on a viscous timescale.

The accretion and cooling timescales (Eqs. (12) and (13)) depend on radius through the radially declining angular velocity $\Omega$ of the torus and the radius $R_{\text{core}}$ at which matter is accreted onto the central star. At a radius of about $54R_\odot$, both timescales are comparable such that cooling is inefficient inside and efficient further out (Extended Data Fig. 1). In our standard model, we therefore assume that the mass interior of $54R_\odot$—that is, the innermost $16.9M_\odot$—transforms into a star with an extended envelope on a viscous timescale. The remaining outer part of the torus is assumed to cool efficiently and evolve into a thin disk. This configuration then forms the initial condition of our 1D stellar evolution computations.

To import the outcome of the 3D simulation into the 1D stellar evolution code MESA, we model a star that has the same chemical and thermal structure as the 3D merger remnant. We first relax a star of given total mass to the chemical structure of the 3D merger remnant before imposing the thermal structure by matching the 3D entropy profile. A comparison of the chemical and entropy structure of the 1D $16.9M_\odot$ merger remnant with the 3D profiles is shown in Extended Data Fig. 4. Our 1D model closely matches the structure of the merger remnant of the 3D simulation.

Setting the rotational profile of the merger remnant requires further consideration. We argued above that the fast viscous evolution of the star-torus structure converts most of the torus into an extended envelope by transporting angular momentum outwards. This angular-momentum transport sets the initial conditions for our 1D merger evolution. In the viscous evolution of the remnant of a double white-dwarf merger, efficient outward angular-momentum transport is found such that the rotational profile of the central star remains a solid-body rotator and smoothly transitions into a near-Keplerian profile at the boundary between the central star and outer disk[45]. The same evolution and outcome has been found by others who studied the aftermath of double-white-dwarf mergers within a prescribed viscosity model but also within more self-consistent MHD simulations[44,46]. In all cases, a large fraction of angular momentum is transported outwards, allowing for the rapid accretion of a large fraction of the torus.

Our 3D merger simulation also shows that the central star reaches solid-body rotation with the angular velocity matching that of the layer between star and torus, which is approximately 80% of the Keplerian value. The surface of the central star does not reach 100% Keplerian rotation because the torus is not only centrifugally supported but also thermally supported. We therefore assume that our merger remnant is a solid-body rotator that rotates at 90% of the critical Keplerian velocity at the surface after the viscous evolution.

*Restructuring of the stellar interior during the thermal relaxation phase after the merger.* During the thermal relaxation, the merger product first approaches critical surface rotation before the model spins down rapidly (Fig. 2 and Extended Data Fig. 5). This spin-down is not driven by angular-momentum loss but can be understood as follows. The internal magnetic field keeps the star close to solid-body rotation such that the total angular momentum $J$ of the star is $J = r_{\text{g}}^2 MR_*^2 \Omega_*$ with the stellar radius $R_*$, the moment of inertia factor $r_{\text{g}}^2$, the stellar mass $M$ and the

angular velocity $\Omega_*$. For constant angular momentum $J$ and mass $M$, the surface rotational velocity evolves according to:

$$v_{rot} = \frac{J}{r_g^2 M R_*} \propto (r_g^2 R_*)^{-1} \qquad (14)$$

In the contraction phase when the star spins down (about $10^2$–$10^4$ yr after the merger), the radius decreases by a factor of 4 while $r_g^2$ increases by a factor of 20, fully explaining the observed spin-down of the merger product by a factor of about 5 (Extended Data Fig. 5). This change in $r_g^2$ is because, after the coalescence, the core of the merger is hotter and denser than in full equilibrium, which leads to core expansion while the envelope contracts (Extended Data Fig. 5). Because magnetic braking is unimportant for the spin-down, our conclusions regarding the final spin and surface properties of the merger product are almost independent of the magnetic field structure and strength used in the 1D stellar models.

## Data availability

The data generated, analysed and presented in this study are available from the corresponding authors on reasonable request.

## Code availability

The AREPO code is publicly available at https://arepo-code.org. The MESA code is publicly available at http://mesa.sourceforge.net.

31. Pakmor, R., Bauer, A. & Springel, V. Magnetohydrodynamics on an unstructured moving grid. *Mon. Not. R. Astron. Soc.* **418**, 1392–1401 (2011).
32. Pakmor, R. & Springel, V. Simulations of magnetic fields in isolated disc galaxies. *Mon. Not. R. Astron. Soc.* **432**, 176–193 (2013).
33. Powell, K. G., Roe, P. L., Linde, T. J., Gombosi, T. I. & De Zeeuw, D. L. A solution-adaptive upwind scheme for ideal magnetohydrodynamics. *J. Comput. Phys.* **154**, 284–309 (1999).
34. Asplund, M., Grevesse, N., Sauval, A. J. & Scott, P. The chemical composition of the Sun. *Annu. Rev. Astron. Astrophys.* **47**, 481–522 (2009).
35. Paxton, B. et al. Modules for Experiments in Stellar Astrophysics (MESA): planets, oscillations, rotation, and massive stars. *Astrophys. J. Suppl. Ser.* **208**, 4 (2013).
36. Paxton, B. et al. Modules for Experiments in Stellar Astrophysics (MESA): binaries, pulsations, and explosions. *Astrophys. J. Suppl. Ser.* **220**, 15 (2015).
37. Paxton, B. et al. Modules for Experiments in Stellar Astrophysics (MESA): convective boundaries, element diffusion, and massive star explosions. *Astrophys. J. Suppl. Ser.* **234**, 34 (2018).
38. Ohlmann, S. T., Röpke, F. K., Pakmor, R. & Springel, V. Constructing stable 3D hydrodynamical models of giant stars. *Astron. Astrophys.* **599**, A5 (2017).
39. Górski, K. M. et al. HEALPix: a framework for high-resolution discretization and fast analysis of data distributed on the sphere. *Astrophys. J.* **622**, 759–771 (2005).
40. Spitzer, L. *Physics of Fully Ionized Gases* 2nd edn (Interscience, 1962).
41. Donati, J.-F. et al. The surprising magnetic topology of $\tau$ Sco: fossil remnant or dynamo output? *Mon. Not. R. Astron. Soc.* **370**, 629–644 (2006).
42. Ud-Doula, A., Owocki, S. P. & Townsend, R. H. D. Dynamical simulations of magnetically channelled line-driven stellar winds—III. Angular momentum loss and rotational spin-down. *Mon. Not. R. Astron. Soc.* **392**, 1022–1033 (2009).
43. Lamers, H. J. G. L. M., Snow, T. P. & Lindholm, D. M. Terminal velocities and the bistability of stellar winds. *Astrophys. J.* **455**, 269 (1995).
44. Shen, K. J., Bildsten, L., Kasen, D. & Quataert, E. The long-term evolution of double white dwarf mergers. *Astrophys. J.* **748**, 35 (2012).
45. Schwab, J., Shen, K. J., Quataert, E., Dan, M. & Rosswog, S. The viscous evolution of white dwarf merger remnants. *Mon. Not. R. Astron. Soc.* **427**, 190–203 (2012).
46. Ji, S. et al. The post-merger magnetized evolution of white dwarf binaries: the double-degenerate channel of sub-Chandrasekhar type Ia supernovae and the formation of magnetized white dwarfs. *Astrophys. J.* **773**, 136 (2013).

**Author contributions** F.R.N.S. initiated the project and carried out the 1D MESA computations. S.T.O. carried out the 3D AREPO simulations. F.R.N.S. and S.T.O. wrote most of the manuscript. P.P. and F.K.R. assisted with the 1D and 3D computations, respectively. S.A.B. in particular helped to analyse and understand the magnetic-field amplification process. V.S. and R.P. wrote the AREPO code and supported S.T.O. with the 3D simulations. All authors contributed to the analysis, discussion and writing of the paper.
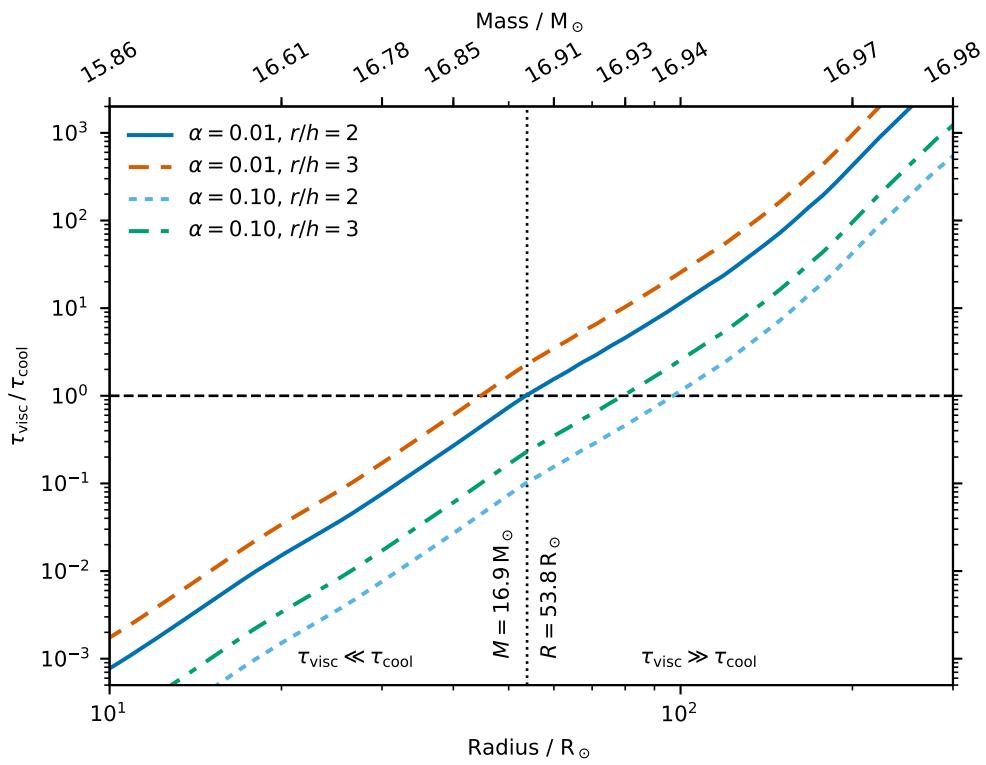
**Extended Data Fig. 1 | Ratio of viscous and cooling timescales.** The $\tau_{\mathrm{visc}}/\tau_{\mathrm{cool}}$ ratio is shown as a function of radius (and mass) of the merger after 6 d for different disk thicknesses $h$ and viscosity parameters $\alpha$.
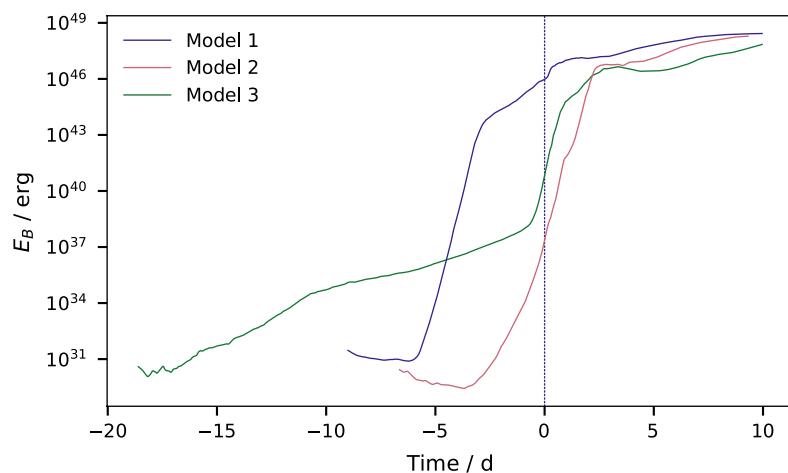
**Extended Data Fig. 2 | Evolution of total magnetic field energy for different simulation setups.** Model 1 is the standard run shown in the main text. Models 2 and 3 have a lower resolution. Model 3 started with a larger initial separation. The times for all models are normalized with the time of merger set to 0.
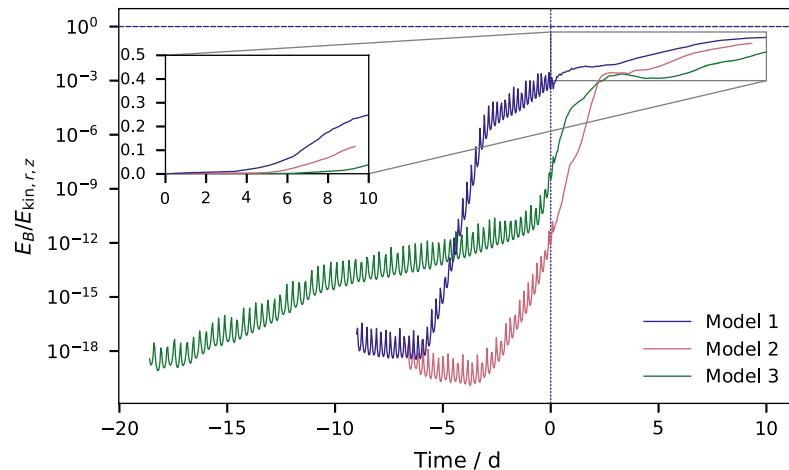
**Extended Data Fig. 3 | Ratio of magnetic and radial, kinetic (that is, turbulent) energy in our 3D MHD simulations.** The magnetic energies, $E_B$, of our models approach equipartition with the turbulent energy, for which we use the kinetic energy of motions in the radial and $z$ directions, $E_{\mathrm{kin},r,z}$, as a proxy. Inset, the ratio of magnetic and kinetic energy on a linear scale, showing that our models reach values of about 5%–30%. The dashed horizontal line indicates equipartition of magnetic and kinetic energy. The small, periodic wiggles in the curves before coalescence are caused by the orbital motion of the binary.
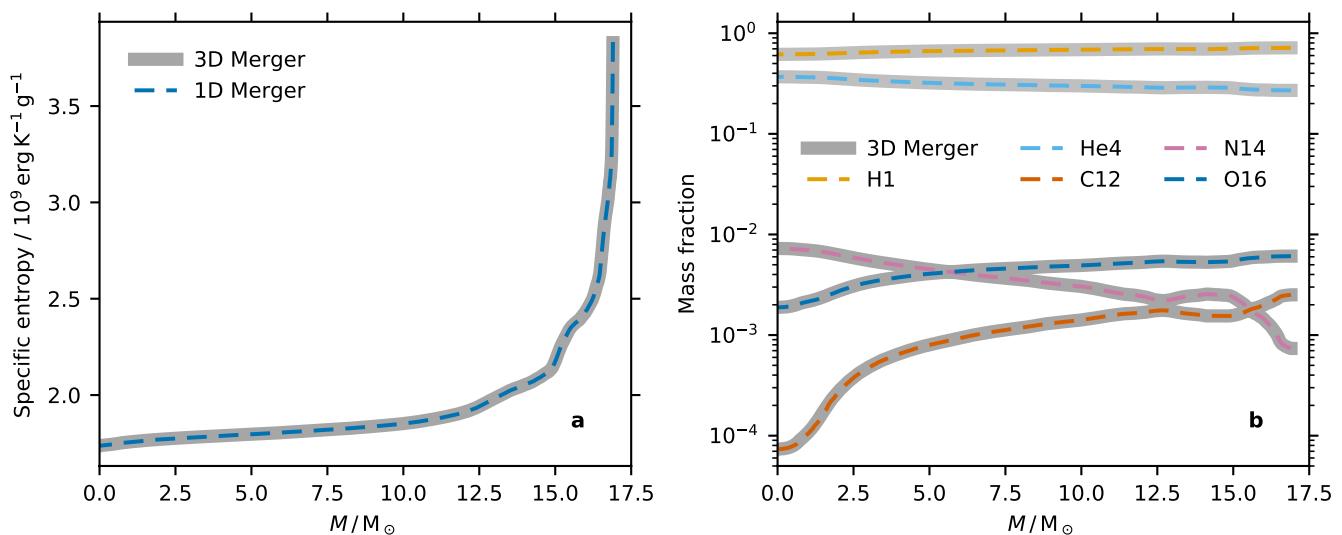
**Extended Data Fig. 4 | Comparison of the final 3D and initial 1D merger product. a, b,** The entropy (**a**) and the hydrogen (H1), helium (He4), carbon (C12), nitrogen (N14) and oxygen (O16) mass fractions (**b**) of the 3D merger remnant (thick grey lines) and the 1D stellar model (dashed lines) are compared.
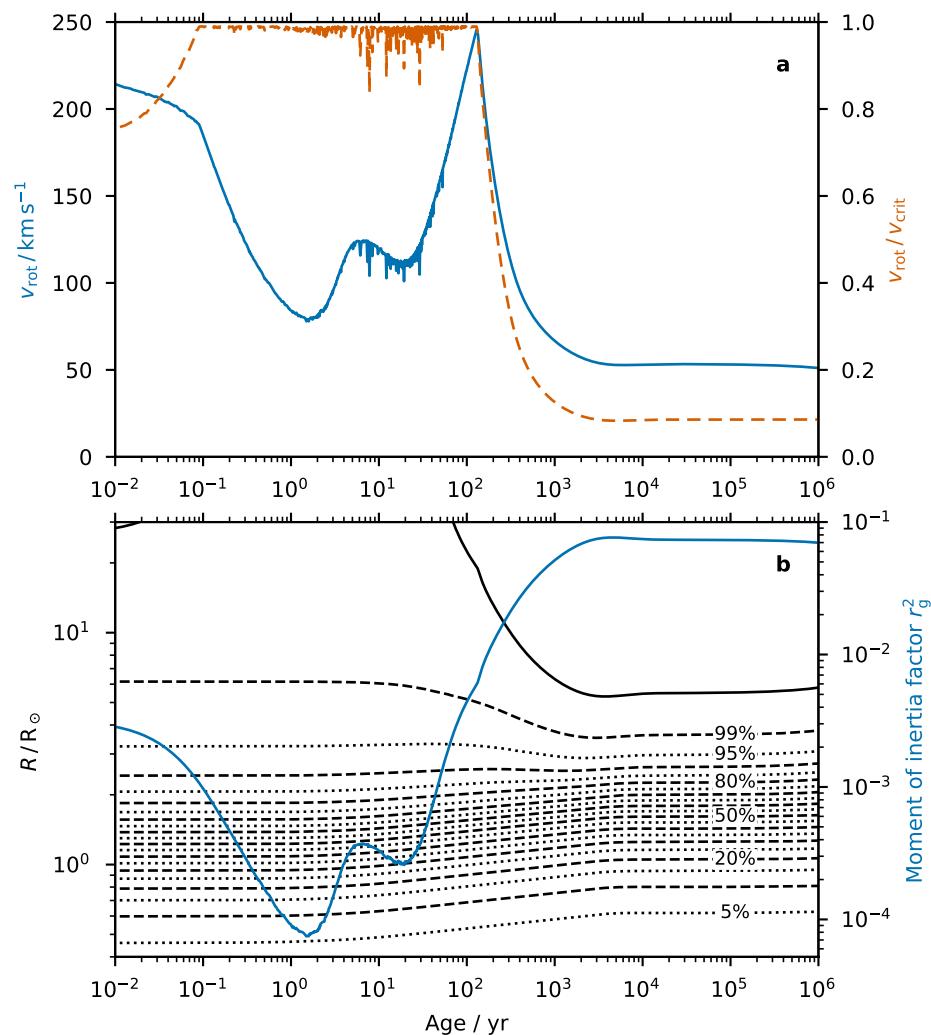
**Extended Data Fig. 5 | Rotational velocity and internal mass readjustment of the 1D merger model. a**, Equatorial surface rotational velocity $v_{rot}$ (blue solid line) and rotational velocity in terms of critical Keplerian velocity $v_{crit}$ (red dashed line) as a function of time after the merger. **b**, Radial location of various mass coordinates in steps of 5% of the total mass (see percentage labels for the black dotted and dashed lines) and moment of inertia factor $r_g^2$ (blue solid line) as a function of time. The black solid line indicates the stellar surface.

# LETTER

# Analogue quantum chemistry simulation

Javier Argüello-Luengo[1,2], Alejandro González-Tudela[1,3]*, Tao Shi[1,4], Peter Zoller[1,5] & J. Ignacio Cirac[1,6]*

**Computing the electronic structure of molecules with high precision is a central challenge in the field of quantum chemistry. Despite the success of approximate methods, tackling this problem exactly with conventional computers remains a formidable task. Several theoretical[1,2] and experimental[3–5] attempts have been made to use quantum computers to solve chemistry problems, with early proof-of-principle realizations done digitally. An appealing alternative to the digital approach is analogue quantum simulation, which does not require a scalable quantum computer and has already been successfully applied to solve condensed matter physics problems[6–8]. However, not all available or planned setups can be used for quantum chemistry problems, because it is not known how to engineer the required Coulomb interactions between them. Here we present an analogue approach to the simulation of quantum chemistry problems that relies on the careful combination of two technologies: ultracold atoms in optical lattices and cavity quantum electrodynamics. In the proposed simulator, fermionic atoms hopping in an optical potential play the role of electrons, additional optical potentials provide the nuclear attraction, and a single-spin excitation in a Mott insulator mediates the electronic Coulomb repulsion with the help of a cavity mode. We determine the operational conditions of the simulator and test it using a simple molecule. Our work opens up the possibility of efficiently computing the electronic structures of molecules with analogue quantum simulation.**

Quantum computers are expected to have a considerable impact in several areas of science because they will be able to tackle problems that are intractable with classical devices. Particularly relevant are quantum many-body problems involving several systems that interact with each other according to the rules of quantum physics[9]. Given the current theoretical and experimental progress, the most timely and important ones are quantum chemistry problems, which generally involve obtaining the ground-state energy of many electrons that interact with nuclei and with each other through Coulomb interactions. Current approaches to the simulation of chemistry problems with quantum computers follow the digital approach[10–12], in which one breaks the complete Hamiltonian into gates that are applied in a time-dependent manner.

An alternative way to address quantum many-body problems is analogue quantum simulation[13]. The idea is to use a well controlled quantum system (the simulator) and engineer its interactions according to the Hamiltonian under investigation. This approach has already been used to address questions that the most advanced classical computers cannot resolve[6–8]. The key feature is that their interactions are either local or short-range, being ideally suited for existing simulators. By contrast, analogue simulation of quantum chemistry requires engineering long-range (Coulomb) interactions between fermionic particles, and no system has been identified so far to fulfil such a requirement. This is why current efforts concentrate in digital simulation.

Here we show how to build an analogue simulator for quantum chemistry problems by bridging two paradigmatic systems, namely, ultracold atoms in optical lattices[14–16] and cavity quantum electrodynamics (QED)[17–21]. Fermionic atoms trapped in a periodic three-dimensional (3D) optical potential play the role of electrons and are subject to additional optical potentials emulating their interaction with

the nuclei. The key feature of the scheme is the trapping of another atomic species in a Mott insulator regime with several internal states such that its spin excitations mediate effective forces between the simulated electrons. We show that even though the interaction is local, one can induce Coulomb-like forces among the fermionic atoms in a scalable manner. Although the setup is discrete and finite, we show that precise results can be obtained for simple molecules with moderate lattice sizes. Apart from the standard advantages of analogue simulation over quantum computing regarding the required control[13], the present scheme does not rely on a judicious choice of molecular orbitals[22], but operates directly in real space, improving convergence to the exact result as the system size increases.

One of the main goals of quantum chemistry is to obtain the low-energy behaviour of $N_e$ electrons and several nuclei when the positions, $\mathbf{r}_n$, of the nuclei are fixed. Using a cubic discretization in real space of $N \times N \times N$ sites, the electronic Hamiltonian contains three terms, $H_{qc} = H_{kin} + H_{nuc} + H_{e-e}$, (using $\hbar = 1$ for the reduced Planck constant, and dropping the spin index)

$$H_{kin} = -t_F \sum_{\langle \mathbf{i}, \mathbf{j} \rangle} f_{\mathbf{i}}^\dagger f_{\mathbf{j}} \tag{1}$$

$$H_{nuc} = -\sum_{n, \mathbf{j}} Z_n V(\|\mathbf{j} - \mathbf{r}_n\|) f_{\mathbf{j}}^\dagger f_{\mathbf{j}} \tag{2}$$

$$H_{e-e} = \sum_{\mathbf{i}, \mathbf{j}} V(\|\mathbf{i} - \mathbf{j}\|) f_{\mathbf{i}}^\dagger f_{\mathbf{i}} f_{\mathbf{j}}^\dagger f_{\mathbf{j}} \tag{3}$$

where $f_{\mathbf{i}}$ are annihilation operators of electrons at site $\mathbf{i}$ fulfilling $\{f_{\mathbf{i}}, f_{\mathbf{j}}^\dagger\} = \delta_{\mathbf{i}, \mathbf{j}}$, and $\langle \mathbf{i}, \mathbf{j} \rangle$ denote nearest-neighbour sites. $H_{kin}$ describes electron hopping at rate $t_F$, $H_{nuc}$ represents the electron–nucleus attraction when the nuclei are at positions $\mathbf{r}_n$, and $H_{e-e}$ accounts for the electron–electron repulsion. In both $H_{nuc}$ and $H_{e-e}$, the attractive/repulsive potential has the standard Coulomb form, $V(r) = V_0/r$. The connection between the length/energy scales of the discrete Hamiltonian $H_{qc}$ and the continuum one is given by:

$$\frac{a_0}{a} = \frac{2t_F}{V_0} \quad \text{and} \quad \text{Ry} = \frac{V_0^2}{4t_F} \tag{4}$$

where $a_0$, $a$ and Ry are the Bohr radius, lattice spacing and Rydberg energy, respectively. Thus, we work in a regime:

$$1 \ll \frac{2t_F}{V_0} \ll \frac{N}{N_e^{1/3}} \tag{5}$$

such that the first inequality prevents discretization effects and the second one guarantees that the molecule fits in the volume of the simulator.

Our simulator then requires three components (see Fig. 1a): (i) cold spin-polarized fermionic atoms hopping in a 3D optical potential with a tunable tunnelling rate, $J_F$, which play the role of electrons[14]. We consider spinless fermions, but the spin degree of freedom can be included using an extra internal level[23] (see Methods). (ii) Additional potentials
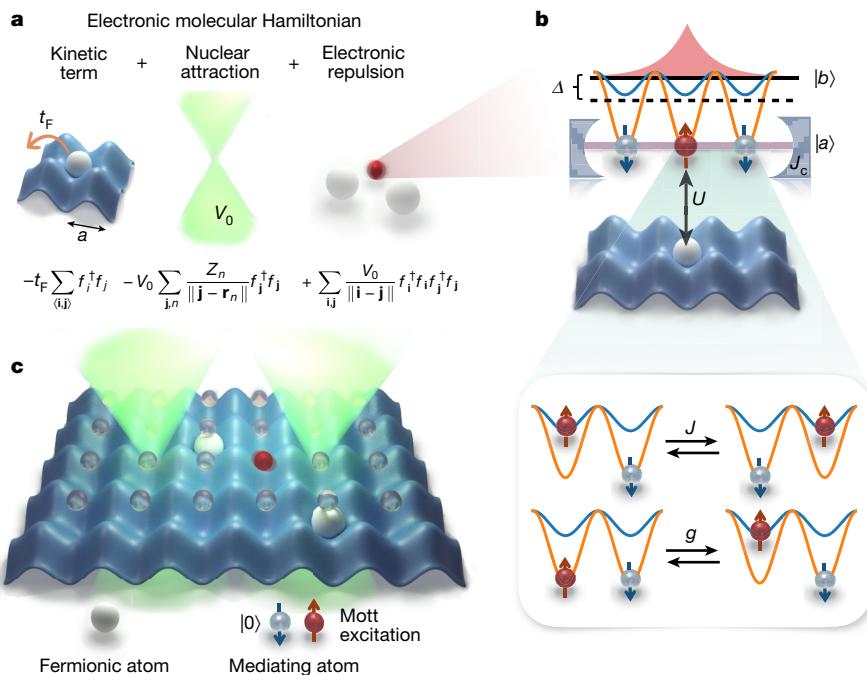
**Fig. 1 | Schematic representation of the analogue simulator.**
**a**, Fermionic atoms, playing the role of electrons, are trapped in a periodic 3D cubic potential. Their hopping simulates the kinetic energy term of the electrons, and they are subject to additional optical potentials that emulate the nuclear interaction. **b**, Coulomb repulsion among the fermions is mediated by a spin excitation of a Mott insulator with three internal levels. Excitations in level $|b\rangle$ are allowed to propagate through spin-exchange

interactions with rate $J$ (inset, top). Excitations in level $|a\rangle$ experience a strong repulsive interaction with the fermions and interact with a cavity mode. These two levels are coupled through either a microwave or a two-photon Raman transition (inset, bottom). **c**, The complete simulator for the $H_2$ molecule. Although a 2D lattice is pictured, the experimental proposal presented here refers to a 3D optical lattice.

to emulate the attraction between fermions and nuclei. Given that this is a single-particle Hamiltonian, it can be created through optical Stark shifts with an adequate spatial modulation. For example, one can use holographic techniques[24,25] with judiciously optimized phase masks to engineer a Coulomb-like spatial potential at the fermionic positions (see Methods). (iii) The most difficult part to simulate is $H_{e-e}$, because it involves repulsive interactions between the fermions with a $1/r$ dependence. Inspired by how virtual photons mediate Coulomb interactions in QED, we use a spin excitation of another atomic species forming a Mott insulator to mediate the Coulomb forces between fermions (see Fig. 1b). This species is composed of $N_M^3$ atoms trapped in an optical potential with the same spacing as that of the fermions and with two additional internal atomic states, $|a\rangle$ and $|b\rangle$, which describe spin excitations. Spin excitations in the $|a\rangle$ state interact repulsively and locally with the fermionic atoms with strength $U$, and propagate through the long-range couplings induced by a cavity mode with rate $J_c$ (refs. [17–21]). The $|b\rangle$ internal state is subject to a different optical potential, such that its itinerant spin excitation propagates through standard nearest-neighbour exchange[26] at rate $J$. Furthermore, an external field (Raman laser or a radiofrequency field) drives the $|a\rangle - |b\rangle$ transition with coupling strength $g$ and detuning $\Delta$. The complete simulator Hamiltonian after the elimination of the cavity mode is $H_{sim} = H_{kin} + H_{nuc} + H_M$, with

$$H_M = \Delta \sum_j b_j^\dagger b_j + J \sum_{\langle i,j\rangle} b_i^\dagger b_j + \frac{J_c}{N_M^3} \sum_{i,j} a_i^\dagger a_j$$
$$+ U \sum_j a_j^\dagger a_j f_j^\dagger f_j + g \sum_j (a_j^\dagger b_j + b_j^\dagger a_j) \qquad (6)$$

where $a_j$ and $b_j$ are annihilation operators for $|a\rangle$ and $|b\rangle$ spin excitations, respectively, at site $j$. Intuitively, the on-site interaction $U$ localizes the $|a\rangle$ and $|b\rangle$ excitations around the fermions, renormalizing their tunnelling rates and creating an effective interaction. Mathematically, one can eliminate the Mott insulator excitations adiabatically and derive the effective dynamics for the fermions (see Supplementary Information



**Fig. 2 | Atomic hydrogen spectrum dependence on the effective Bohr radius. a**, Lower part of the spectrum of the atomic Hamiltonian $H_{qc}$ for a cubic lattice of $N = 100$. Different symbols represent different energy orbitals, and the first three atomic levels ($n = 1, 2, 3$) of the continuum Hamiltonian are represented by the dashed lines. In the blue-shaded region, the hopping parameter is $t_F/V_0 < 0.5$, the Bohr radius is smaller than the lattice spacing, and energies are highly affected by the cut-off of the nuclear potential. As the hopping parameter $t_F/V_0$ increases, the simulator effectively zooms in on the system, as we show in the bottom insets, which present the fermionic density of the second-lowest energy orbitals shown in the main graph. By increasing $t_F/V_0$ we include more lattice sites in the simulation, reducing systematic deviations as $(t_F/V_0)^{-2}$ (red dashed line in top inset), as we show in the top inset for the lowest energy state (see Supplementary Information section 1). At higher values of $t_F/V_0$, solutions suffer from finite-size effects. **b**, Axial cut in the central positions of the lattice for the first nine eigenstates of $H_{qc}$ for $t_F/V_0 = 2$ and $N = 150$. **c**, By choosing the appropriate Bohr radius, the same orbital can be obtained with $N = 1,000$ (top; $t_F/V_0 = 150$) or $N = 20$ (bottom; $t_F/V_0 = 3$), where the discretization of the system is more noticeable.
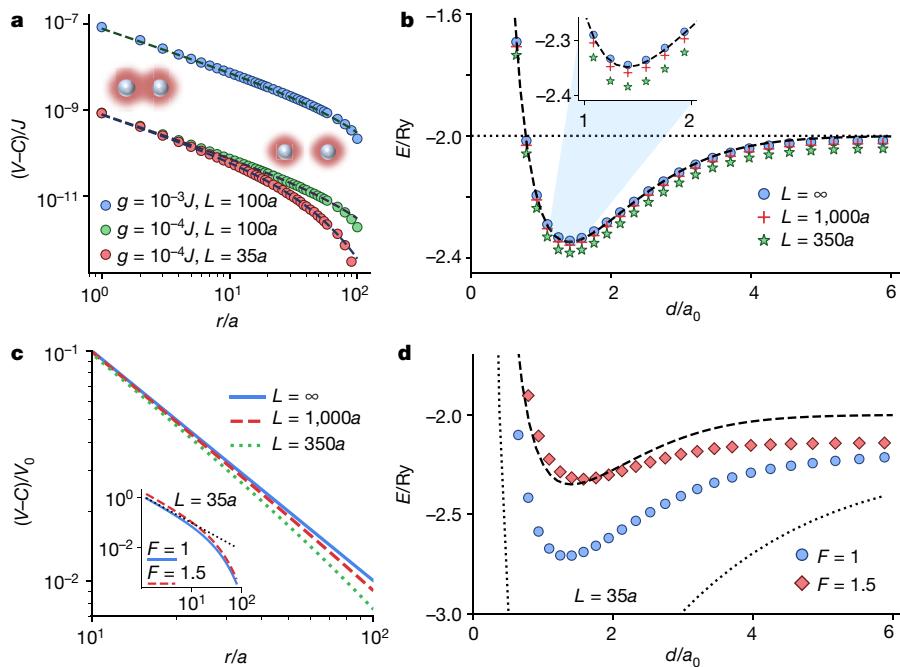
**Fig. 3 | Molecular potential and effective interaction mediated by the Mott insulator. a**, Energy of the single-excitation bound state of the Hamiltonian of equation (6) for two fixed fermions as a function of their separation, $r$. We choose $\Delta = 2J$, $N_M = 200$ and $J_c = J$ to satisfy conditions (5), (8)–(11). The Yukawa potential of equation (7) corresponding to each configuration of parameters is plotted with dashed lines. **b**, We use this effective interaction to calculate the molecular potential associated with an analogue simulator of $H_2$ of size $N = 75$. For each internuclear separation, we choose the $t_F/V_0$ value to give optimal accuracy (see Supplementary Information section 3), ranging from $t_F/V_0 = 4.2$ to $t_F/V_0 = 2.3$ at the dissociation limit (dotted line). Molecular orbitals are included

in the projective basis until convergence is observed. For a Coulomb potential (blue dots), the result agrees with an accurate solution in the nonrelativistic regime[28,29] (dashed line). As $L$ decreases, the exponential decay in the Yukawa potential prevails, underestimating Coulomb repulsion and lowering the molecular potential. **c**, This underestimation of the repulsive potential is more evident when the condition $N \ll L/a$ is violated (inset). **d**, By changing the ratio $F$ between the electronic and nuclear potentials, one can explore artificial repulsive interactions that form pseudomolecules in more relaxed experimental conditions. The dotted line represents the limit of zero repulsion in the absence of a mediating excitation.

section 2). The fermionic part of the simulator Hamiltonian is $H_{sym} \approx H_{qc}$, with $t_F = J_F(N_e - 1)/N_e$, where the electron–electron potential follows a Yukawa form[27] with a constant energy shift $C$:

$$V(r) \approx C + \frac{V_0}{r/a} e^{-r/L} \qquad (7)$$

where $L/a = \sqrt{J/(U + \Delta + \rho_M J_c - 6J)}$ is the localization length, which can be tuned with $\Delta$, $V_0 = g^2/(2\pi J N_e)$ is the strength of the potential repulsion, and $\rho_M = N_e/N_M^3$. This mapping between $H_{sym}$ and $H_{qc}$ holds as long as

$$J_c \ll U \qquad (8)$$

$$J_F \ll J_c \rho_M \quad \text{and} \quad V_0 \sqrt{N_e} \ll J_c \rho_M \qquad (9)$$

$$V_0 N_e^{7/3} \ll J(aN/L)^2 \qquad (10)$$

Condition (8) enforces that the $|a\rangle$ excitation is localized symmetrically only around the position of the fermions; condition (9) guarantees that neither the tunnelling of the fermions nor the interaction with the $|b\rangle$ excitations dominate over the cavity interaction; and condition (10) ensures that the Yukawa potential does not depend on the fermionic positions. Furthermore, to obtain a truly Coulomb repulsion, the length $L$ must be larger than the fermionic lattice with $N$ sites but smaller than the Mott insulator size, that is:

$$N \ll L/a < N_M \qquad (11)$$

When inequalities (5), (8)–(11) are satisfied, the exact solution in the continuum limit is recovered in the limit $N_M > N \to \infty$. Thus, the

finite size of the simulator is what ultimately limits the precision of the simulation.

We now benchmark our simulator for moderate system sizes using numerical simulations. In Fig. 2 we solve the hydrogen problem in a lattice to explore discretization and finite-size effects by comparing the energies of the low-lying excited states with that of the continuum. We show that an error of 0.3% with respect to the exact energy can be obtained for systems of $N = 100$. In Fig. 3 we analyse the accuracy of the simulator for the simplest molecule, $H_2$. First, we compute exactly the energy of the spin excitation that mediates the fermionic repulsion, as a function of the interfermionic separation (Fig. 3a). We show that it reproduces the $1/r$ behaviour over a wide range of values of $g/J$ and $L$. In Fig. 3b we present the molecular potential for $N = 75$ computed using a Yukawa electronic potential with different lengths $L$. We observe excellent qualitative agreement for all $L$ values considered in the figure, and quantitative matching when $L \gtrsim aN$. Remarkably, even if $L \lesssim aN$, valuable information can still be extracted by adjusting other experimental parameters. In Fig. 3d we illustrate how one can increase $V_0$ of the electron repulsion to compensate for the underestimation of the potential at long distances, and obtain a pseudomolecular potential that is qualitatively similar to the one expected with Coulomb interactions.

We consider some practical issues for the implementation of these ideas (see Methods). As atomic species fulfilling the requirements we propose to use two isotopes of alkaline-earth atoms, $^{87}Sr$ and $^{88}Sr$, for the fermions and bosons, respectively. The quantum simulator can be initialized using adiabatic preparation, where the hopping, nuclear attraction and interactions are sequentially turned on. We also propose to read out the physical properties of the system by measuring its total energy through detection of the kinetic energy of the fermionic atoms under different conditions and repeating the measurements. In Methods we discuss some sources of errors and ways to circumvent
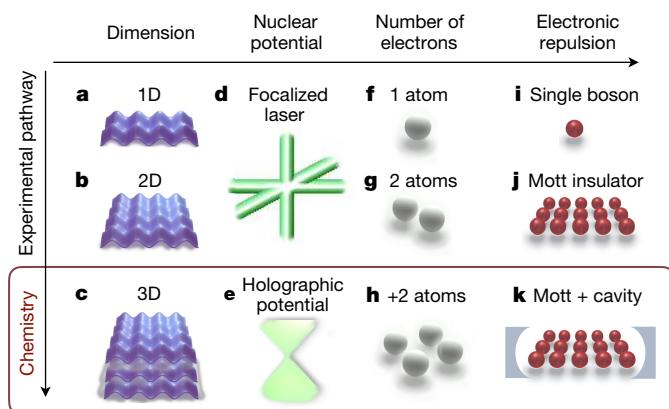
**Fig. 4 | Experimental pathway. a–k**, Schematic representation of the simplifications that can be considered for the different interactions of the system. The lowest row (**c**, **e**, **h**, **k**) corresponds to our full proposal and chemical interactions that we observe in nature. Chemistry in lower dimensions could be considered by restricting the optical lattice accordingly (**a**, **b**). The holographically created Coulomb potential could be replaced by, for example, the Gaussian profile of a focalized laser (**d**), giving a different scaling for the electron–nucleus attraction. First implementations with single atoms (**f**) would allow to observe simple electronic orbitals (such as the energy levels of hydrogen), with no need to mediate repulsion. For only two atoms (**g**) no symmetrizing effect of the cavity is needed when mediating the Coulomb repulsion with the Mott insulator (**j**). Different scalings for repulsion can also be explored in more simplified setups, such as those using a single boson that hops on the lattice and interacts on site with the fermionic atoms (**i**).

them. We emphasize that some of the elements and conditions required in this approach are beyond the capabilities of the current experimental state of the art. However, the rapid progress of analogue quantum simulation may lead to the realization of the present ideas in the near future, motivated by its potential impact in the determination of chemical structures, the understanding of reaction mechanisms or the development of molecular electronics. Furthermore, with judicious changes in its implementation (for example, different optical potential geometries), conditions (8)–(10) may be relaxed. Thus, we believe that our method is a promising complementary approach to fault-tolerant quantum computers.

In summary, we have shown how to simulate quantum chemistry problems using cold atoms in optical lattices embedded in a cavity. We expect that the proposed technique will stimulate both theoretical and experimental research, even before the realization of a fully fledged analogue simulator for quantum chemistry. Figure 4 provides a roadmap of experiments with increasing complexity (from top left to bottom right) towards complete analogue chemistry simulation. For instance, the first experiments could be performed with spinless fermions in one (1D) and two (2D) spatial dimensions. Another simplification might come from non-Coulomb nuclear potentials—for example, in the form of a Gaussian, which does not require holographic techniques—or by using simpler schemes to obtain the fermionic potential—for example, using a single boson without a cavity instead of a Mott insulator. The latter still provides effective repulsion between two fermions. With these simplifications, there is a clear pathway from state-of-the-art setups towards more challenging experimental setups that is based on technological progress. Most importantly, in all these intermediate proof-of-principle setups one could already observe molecular-like potentials, dissociation and other fundamental phenomena in chemistry. Besides, such experiments can also prove to be valuable for benchmarking various numerical techniques and trigger the development of new theoretical methods, thus reaching a deeper understanding of chemistry problems that are challenging to test with classical computers.

## Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-019-1614-4.

1. Aspuru-Guzik, A., Dutoi, A. D., Love, P. J. & Head-Gordon, M. Simulated quantum computation of molecular energies. *Science* **309**, 1704–1707 (2005).
2. Wecker, D., Bauer, B., Clark, B. K., Hastings, M. B. & Troyer, M. Gate-count estimates for performing quantum chemistry on small quantum computers. *Phys. Rev. A* **90**, 022305 (2014).
3. O'Malley, P. et al. Scalable quantum simulation of molecular energies. *Phys. Rev. X* **6**, 031007 (2016).
4. Kandala, A. et al. Hardware-efficient variational quantum eigensolver for small molecules and quantum magnets. *Nature* **549**, 242–246 (2017).
5. Lanyon, B. P. et al. Towards quantum chemistry on a quantum computer. *Nat. Chem.* **2**, 106–111 (2010).
6. Choi, J.-y. et al. Exploring the many-body localization transition in two dimensions. *Science* **352**, 1547–1552 (2016).
7. Trotzky, S. et al. Probing the relaxation towards equilibrium in an isolated strongly correlated one-dimensional Bose gas. *Nat. Phys.* **8**, 325–330 (2012).
8. Greiner, M., Mandel, O., Esslinger, T., Hänsch, T. W. & Bloch, I. Quantum phase transition from a superfluid to a Mott insulator in a gas of ultracold atoms. *Nature* **415**, 39–44 (2002).
9. Feynman, R. P. Simulating physics with computers. *Int. J. Theor. Phys.* **217**, 467–488 (1982).
10. Cao, Y. et al. Quantum chemistry in the age of quantum computing. *Chem. Rev.* https://doi.org/10.1021/acs.chemrev.8b00803 (2019).
11. Kassal, I., Whitfield, J. D., Perdomo-Ortiz, A., Yung, M.-H. & Aspuru-Guzik, A. Simulating chemistry using quantum computers. *Annu. Rev. Phys. Chem.* **62**, 185–207 (2011).
12. Reiher, M., Wiebe, N., Svore, K. M., Wecker, D. & Troyer, M. Elucidating reaction mechanisms on quantum computers. *Proc. Natl Acad. Sci. USA* **114**, 7555–7560 (2017).
13. Cirac, J. I. & Zoller, P. Goals and opportunities in quantum simulation. *Nat. Phys.* **8**, 264–266 (2012).
14. Bloch, I., Dalibard, J. & Zwerger, W. Many-body physics with ultracold gases. *Rev. Mod. Phys.* **80**, 885–964 (2008).
15. Murch, K. W., Moore, K. L., Gupta, S. & Stamper-Kurn, D. M. Observation of quantum-measurement backaction with an ultracold atomic gas. *Nat. Phys.* **4**, 561–564 (2008).
16. Krinner, L., Stewart, M., Pazmiño, A., Kwon, J. & Schneble, D. Spontaneous emission of matter waves from a tunable open quantum system. *Nature* **559**, 589–592 (2018).
17. Ritsch, H., Domokos, P., Brennecke, F. & Esslinger, T. Cold atoms in cavity-generated dynamical optical potentials. *Rev. Mod. Phys.* **85**, 553–601 (2013).
18. Schreppler, S. et al. Optically measuring force near the standard quantum limit. *Science* **344**, 1486–1489 (2014).
19. Brennecke, F., Ritter, S., Donner, T. & Esslinger, T. Cavity optomechanics with a Bose–Einstein condensate. *Science* **322**, 235–238 (2008).
20. Brennecke, F. et al. Cavity QED with a Bose–Einstein condensate. *Nature* **450**, 268–271 (2007).
21. Domokos, P. & Ritsch, H. Collective cooling and self-organization of atoms in a cavity. *Phys. Rev. Lett.* **89**, 253003 (2002).
22. Szabo, A. & Ostlund, N. S. *Modern Quantum Chemistry: Introduction to Advanced Electronic Structure Theory* (Courier Corporation, 2012).
23. Daley, A. J., Boyd, M. M., Ye, J. & Zoller, P. Quantum computing with alkaline-earth-metal atoms. *Phys. Rev. Lett.* **101**, 170504 (2008).
24. Sturm, M. R., Schlosser, M., Walser, R. & Birkl, G. Quantum simulators by design: many-body physics in reconfigurable arrays of tunnel-coupled traps. *Phys. Rev. A* **95**, 063625 (2017).
25. Barredo, D., Lienhard, V., de Léséleuc, S., Lahaye, T. & Browaeys, A. Synthetic three-dimensional atomic structures assembled atom by atom. *Nature* **561**, 79–82 (2018).
26. Riegger, L. et al. Localized magnetic moments with tunable spin exchange in a gas of ultracold fermions. *Phys. Rev. Lett.* **120**, 143601 (2018).
27. de Vega, I., Porras, D. & Cirac, J. I. Matter-wave emission in optical lattices: single particle and collective effects. *Phys. Rev. Lett.* **101**, 260404 (2008).
28. Kołos, W. & Wolniewicz, L. Accurate adiabatic treatment of the ground state of the hydrogen molecule. *J. Chem. Phys.* **41**, 3663–3673 (1964).
29. Sims, J. S. & Hagstrom, S. A. High precision variational calculations for the Born–Oppenheimer energies of the ground state of the hydrogen molecule. *J. Chem. Phys.* **124**, 094101 (2006).

## METHODS

Here we provide further details and candidates for the experimental implementation of some of the key components of the proposed method.

**Holographic engineering of the nuclear potential.** Holographic techniques are a possible approach to engineering the 3D Coulomb potential 'seen' by the fermionic atoms, $H_{nuc}$. Following this method, the potential is created by imprinting a phase pattern on an incoming laser, as in ref. [25], where it was used to experimentally induce microtrap potentials for single atoms. The same technique can be applied to engineer various intensity patterns at the lattice sites where the fermions are trapped by choosing an appropriate phase mask. To do this, we particularized the algorithm used in that implementation (Gerchberg–Saxton algorithm[30], G-S) to identify the phase mask that provides the required geometry (a 3D Coulomb potential, $V_j = V_0/||\mathbf{j} - \mathbf{n}||$).

We then use the 'ping-pong' strategy described by the discretized G-S algorithm to retrieve the phases of the holographic field $u$ associated with our intensity distribution ($|u_j|^2 = V_j$). This approach iterates over the following steps[31]: (i) the field is transformed to reciprocal space via a fast Fourier transform (FFT), obtaining $\tilde{u} = \text{FFT}(u)$. (ii) We restrict the calculation to those $\mathbf{k}$ vectors that satisfy the physical constraints of the system; that is, they must sit on an Ewald sphere of radius $k_0 = 2\pi/\lambda$, where $\lambda$ is the wavelength of the monochromatic input light. The remaining contributions are neglected, defining a constrained field in Fourier space, $\tilde{u}^c$. (iii) We use an inverse FFT (IFFT) to transform the field back to real space, $u^c = \text{IFFT}(\tilde{u}^c)$. At this point, $u^c$ satisfies diffraction laws but may differ from the goal potential. (iv) The objective potential is combined with the phases of the constrained solution, $\varphi_j = \arg(u_j^c)$, giving $u_j = \sqrt{V_j} e^{i\varphi_j}$. To improve the accuracy, we use a refined phase mask of size $(n_{div}N) \times (n_{div}N)$, with $N$ being the number of lattice sites per side and $n_{div}$ a refining factor (see insets in Extended Data Fig. 1). Because the trapped atoms are only affected by the value of the holographic potential at the lattice sites, these are the only coordinates where the field is updated in step (iv).

In Extended Data Fig. 1 we show the result of applying this optimization process to the Coulomb potential. By quantifying the normalized relative error[32]

$$\varepsilon = \frac{1}{N^3} \sum_j \frac{\left(\sqrt{V_j^*} - \sqrt{V_j}\right)^2}{|V_j|} \quad (12)$$

between the desired Coulomb potential ($V_j$) and the intensity pattern obtained numerically ($V_j^* = |u_j|^2$), we observe that the accuracy obtained for $n_{div} = 3$ already provides normalized relative errors smaller than 0.3% for $N = 30$. More precise results can be obtained by increasing the number of iterations of the algorithm or the refining factor.

**Candidate atomic species.** The proposed method is based on the interplay between two atomic species: (i) fermionic atoms, which have two internal levels and play the role of the electronic spin; and (ii) mediating atoms, which must have three levels available: $|0\rangle$ for the ground state in the Mott insulator, $|a\rangle$ for the spin excitation and $|b\rangle$ for the state that tunnels and induces the effective repulsion.

All atoms need to see the same lattice, although with tunable tunnelling amplitudes. The fermions must additionally see the external potential and have to interact with the internal state of the mediating atoms. The scattering lengths for the interactions corresponding to levels $|0\rangle$, $|a\rangle$ and $|b\rangle$ do not need to be the same, and we require the scattering length between the fermions and the mediating atom in state $|b\rangle$ to be negligible, so that the scaling of the moving excitation is unchanged. Additionally, the mediating atoms in $|a\rangle$ have to be exposed to the cavity mode. We note that the atoms that mediate the fermionic interactions can correspond to both bosons and fermions, as there is only a single excitation.

Over the past years, many atomic species have been trapped, condensed and used in experiments with optical lattices. For illustration, let us give a particular example using fermionic and bosonic alkaline-earth atoms. These atoms offer a rich internal structure, with long-lived excited metastable states $^3P_0$ and $^3P_2$.

We can use $^{87}$Sr as a simulator for electrons, which has a nuclear spin of $I = 9/2$. Similarly to ref. [23], one can encode the spin of the simulated electron into the nuclear states $|\uparrow\rangle \equiv |^1S_0, m_I = -9/2\rangle$ and $|\uparrow\rangle \equiv |^1S_0, m_I = -7/2\rangle$. This information is therefore protected from the electronic transitions used in the rest of the process.

One can now use one of the bosonic isotopes of Sr, $^{88}$Sr, for the mediating atoms. We assign the long-lived states $^1S_0$, $^3P_0$ and $^3P_2$ as levels $|0\rangle$, $|a\rangle$ and $|b\rangle$, respectively. Because there is a 'magic' wavelength at which the trapping potential for the S state is equal to that of one of the P states, it is possible to choose nearby frequencies that provide essentially the same lattice period for all states with the appropriate conditions. We note that the isotope shift in these atoms is sufficiently large to induce the external potential of the fermionic isotope without affecting the mediating species. This can be done by using an additional laser driving the $^1S_0$ level off resonance (with the holographic techniques explained above). The cavity can also

be tuned close to the $^3P_0 \to {}^3S_1$ resonance without affecting the other states. Two-photon Raman transitions can then couple levels $^3P_0$ and $^3P_2$ appropriately, or one can also use a four-photon transition with level $^1S_0$. We also note that having nuclear spin 0 is not a problem in this case, because the mediated electronic repulsion is spin-independent. By choosing an isotope of Sr, an optical lattice with the same spacing can be engineered at the bosonic state $^1S_0$, similarly to the fermionic atoms. Furthermore, the lattice depth and spacing in $^3P_2$ can be independently controlled (for example, by using magic wavelengths[33]) and its scattering length can be tuned (for example, using non-resonant light[34]).

**Adiabatic preparation.** A crucial step is the preparation of the appropriate states of the simulator, which needs to be carefully designed for the molecular system that one wants to explore. In general, one is interested in the ground-state properties of the electronic configuration for given positions of the nuclei. In this line, one option consists of sequentially adding the different interactions involved in the simulation.

As an example, we illustrate this approach for the ground state of $H_2$, which corresponds to a singlet; that is, electrons have opposite spins and are therefore distinguishable. We propose the following strategy. First, we create a product state of tightly trapped fermions that respects the symmetries of the problem. Secondly, we slowly allow fermions to hop; then, the attraction to the nuclear potential leads to single-particle orbitals. Finally, we adiabatically introduce the electronic repulsion, completing the Hamiltonian.

As an initial state, we begin with both fermions tightly trapped in a lattice site $\mathbf{j}_0$ ($t_F(0) = 0$). Using an external laser we create a single Mott excitation at level $a$ at this position. The initial on-site interaction $U \neq 0$ with the fermionic atoms maintains this excitation at $\mathbf{j}_0$. We can define this state as $|\psi_0\rangle = f_{\uparrow,\mathbf{j}_0}^\dagger f_{\downarrow,\mathbf{j}_0}^\dagger a_{\mathbf{j}_0}^\dagger |0\rangle$, where $f_{\uparrow,\mathbf{n}}^\dagger$ ($f_{\downarrow,\mathbf{n}}^\dagger$) denotes the creation operator of one spin-up (spin-down) fermionic atom at position $\mathbf{n}$, and $a_{\mathbf{n}}^\dagger$ the creation of a single excitation of the Mott insulator at that site. One can now adiabatically evolve this initial state according to the following steps.

(I) Both the fermionic atoms and the Mott excitations are coherently transported into a symmetric superposition of two positions, $\mathbf{n}_1$ and $\mathbf{n}_2$, where the nuclear potentials will be centred, separated by the desired number of sites, $d/a$. That is, a new state $|\psi_I\rangle = 2^{-3/2}(f_{\uparrow,\mathbf{n}_1}^\dagger + f_{\uparrow,\mathbf{n}_2}^\dagger)(f_{\downarrow,\mathbf{n}_1}^\dagger + f_{\downarrow,\mathbf{n}_2}^\dagger)(a_{\mathbf{n}_1}^\dagger + a_{\mathbf{n}_2}^\dagger)|0\rangle$. One can use different strategies to do this, such as using a moving double-well potential that adiabatically transports the wavepacket in opposite directions[35] or standing waves[36]. As the on-site interaction ($U$) is present, the Mott excitation is also transported, mediated by the long-range cavity interaction $J_c$.

At this point, the holographic Coulomb potential described in the previous sections is applied. Given that hopping processes are inhibited and fermionic atoms are now already centred in the nuclear positions, no evolution will be observed. Also, the coupling between the excited levels $a$ and $b$ is switched off ($g = 0$). Thus, the repulsive interaction mediated by the Mott insulator is inactive. The on-site interaction $U$ modifies the fermionic hopping ($J_F \to t_F$) as detailed in the main text. Given that the optical lattice is infinitely deep ($t_F = 0$) at this point, it translates into an effective null Bohr radius ($a_0(0)/a = 0$). The nuclear separation is then infinite ($d/a_0 = \infty$), and we have prepared the dissociation limit.

(II) The next step consists of increasing the orbital size. For this, we adiabatically relax the optical lattice by slowly increasing the value of $t_F$. From the point of view of the molecular potential, this corresponds to growing the Bohr radius and therefore decreasing the effective distance $d/a_0$ (even though $d/a$ remains fixed). As only the kinetic and attractive terms of the Hamiltonian are present, there is no interaction between the two fermions, and the resulting eigenstate corresponds to two independent ground-state electrons associated with $H_2^+$.

(III) Once these single-particle orbitals are prepared, one can adiabatically introduce the repulsive interactions. For this, we slowly couple Mott levels $|a\rangle$ and $|b\rangle$ using a Raman transition of intensity $0 \to g$. As discussed in the main text, this is the final ingredient required to induce an effective repulsion proportional to $g^2$ among the fermionic atoms. One can then increase $g$ until the repulsion equals $V_0$.

In Extended Data Table 1 we summarize the parameters that are modified at each stage. In Extended Data Fig. 2 we show the numerical simulations of this adiabatic path in two dimensions (and $1/r$ interactions) to prepare the ground state for a nuclear separation of $d/a_0 = 10$, demonstrating that such adiabatic preparation is indeed feasible within reasonable parameters.

**Measurement.** From the chemistry perspective, all relevant quantities can be expressed in terms of the fermionic density. This is the approach used, for example, in density functional theory methods[37]. One possibility then is to perform a 3D spatial tomography of the $N_e$ electrons and reconstruct the fermionic density. Although this is very complex in practice, it could be achieved with gas microscopy techniques (see ref. [38], for instance).

An alternative approach would be to measure the energy of the system. In addition to constructing molecular potentials, scanning the energy at different nuclear configurations can provide additional information, such as the magnitude of the molecular forces (Hellmann–Feynman theorem[22]). For this, three quantities need

to be measured simultaneously: the expected values of the kinetic energy, $\langle K \rangle$, the nuclear attraction, $\langle V_n \rangle$, and the electronic repulsion, $\langle V_e \rangle$, such that the total energy is $E = \langle K \rangle + \langle V_n \rangle + \langle V_e \rangle$. By using sudden quenches of the Hamiltonian, such contributions can be independently converted into kinetic energy. One can then perform a time-of-flight measurement of the fermionic atoms expelled from the lattice using, for example, ionization or fluorescence techniques. As one can observe, the measured quantities will not correspond to eigenstates of the original Hamiltonian, which will introduce a variance proportional to the number of fermions. One could then repeat this procedure to gain statistical significance. Once the equilibrium point of the molecular potential is identified, the procedure can be greatly simplified, because only the $\langle K \rangle$ measurement is needed to read the total energy at that point, according to the virial theorem for molecules[39].

**Experimental considerations.** A reliable simulation of the quantum chemistry Hamiltonian requires that our simulator, described by equation (6), satisfies inequalities (5), (8)–(11). We are, however, aware that there will be other experimental imperfections that may impose extra conditions and that will have to be analysed in detail to optimize the performance of the simulation. The most relevant ones are:

(1) Finite temperature leads to thermal fluctuations, which may spoil the simulation by populating undesirable higher-energy bands. Thus, these fluctuations will lead to defects in the Mott insulator (see below) and may also influence the internal states of the atoms. The latter influence, however, can be very well controlled in atomic systems, given that we only need the atoms to be initially in a polarized state, which is reasonably easy to prepare.

(2) Dephasing can be initiated by fluctuations in the transitions or by magnetic fields (as internal levels are being used). This would limit the potential of the system as a quantum simulator. However, the first effect is small in the case of microwave or Raman transitions, and the second one can be controlled under the conditions used for condensed matter simulations[9,10].

(3) Inexact fermionic filling. Because fermions play the role of electrons, an inexact number of fermionic atoms hopping in the lattice translates into an erroneous effective charge in the simulated molecule. These errors can be post-selected by measuring the number of fermionic atoms after the simulation is performed.

(4) Defects in the Mott insulator. The absence of Mott particles in a given lattice site will locally modify the effective fermion potential. Fermions hopping to this site cannot mediate its repulsive interaction through spin excitations, perturbing the simulated molecular orbital around this position. Importantly, such defects will not affect the potential far from the fermion so that the final performance of the simulation will scale with the density of defects instead of with their number.

(5) Spatial inhomogeneities in cavity coupling. In the simulator Hamiltonian of equation (6), we assumed that the $|a\rangle$ excitations couple homogeneously to the cavity mode. In general, there might be some inhomogeneities that translate into a Hamiltonian:

$$\frac{J_c}{N_M^3} \sum_{\mathbf{i},\mathbf{j}} f_{\mathbf{i},\mathbf{j}} a_\mathbf{i}^\dagger a_\mathbf{j} \qquad (13)$$

The fluctuations of $f_{\mathbf{i},\mathbf{j}}$ around 1 will induce an extra decoherence time of $\Gamma_{c,inh}$, which must also be smaller than our simulator parameters. In state-of-the-art experiments, optical cavities at wavelengths of 780 nm and with a beam waist of around 60 μm are already available[40], which would roughly allow 50–100 local minima of the standing wave to sit in a homogeneous region.

(6) Cavity and atom losses. Even though the cavity-mediated interactions are mediated by a virtual population of photons, the cavity decay introduces extra decoherence into the system due to the emission of these virtual photons. Moreover, the atomic excited states, which are also virtually populated, may decay to free space, introducing losses. Thus, the cooperativity of the cavity QED system must be large to avoid both type of losses.

(7) Three-body losses. Because we have fermions and there can be at most one atom per lattice site, this type of loss should be small.

Therefore, most of the possible errors of the simulation are either already under control in current experiments[9,10] or depend only on the number of defects.

## Code and data availability

The computer code developed for this study is available from the corresponding authors upon reasonable request. All data supporting the findings of this study can be generated using the numerical methods described within Methods and Supplementary Information and are available upon reasonable request.

30. Gerchberg, R. & Saxton, O. A practical algorithm for the determination of phase from image and diffraction plane pictures. *Optik* **35**, 237–246 (1971).
31. Whyte, G. & Courtial, J. Experimental demonstration of holographic three-dimensional light shaping using a Gerchberg–Saxton algorithm. *New J. Phys.* **7**, 117 (2005).
32. Soifer, V. A., Kotlar, V. & Doskolovich, L. *Iterative Methods for Diffractive Optical Elements Computation* (CRC Press, 2014).
33. Ye, J., Kimble, H. & Katori, H. Quantum state engineering and precision metrology using state-insensitive light traps. *Science* **320**, 1734–1738 (2008).
34. Crubellier, A., González-Férez, R., Koch, C. P. & Luc-Koenig, E. Controlling the s-wave scattering length with nonresonant light: predictions of an asymptotic model. *Phys. Rev. A* **95**, 023405 (2017).
35. Mandel, O. et al. Coherent transport of neutral atoms in spin-dependent optical lattice potentials. *Phys. Rev. Lett.* **91**, 010407 (2003).
36. Anderlini, M., Sebby-Strabley, J., Kruse, J., Porto, J. V. & Phillips, W. D. Controlled atom dynamics in a double-well optical lattice. *J. Phys. B* **39**, S199–S210 (2006).
37. Parr, R. G. & Yang, W. *Density-Functional Theory of Atoms and Molecules* (Oxford Univ. Press, 1989).
38. Omran, A. et al. Microscopic observation of Pauli blocking in degenerate fermionic lattice gases. *Phys. Rev. Lett.* **115**, 263001 (2015).
39. Gupta, V. P. *Principles and Applications of Quantum Chemistry* (Academic Press, 2015).
40. Schleier-Smith, M. H., Leroux, I. D., Zhang, H., Van Camp, M. A. & Vuletić, V. Optomechanical cavity cooling of an atomic ensemble. *Phys. Rev. Lett.* **107**, 143005 (2011).

**Author contributions** J.A.-L., A.G.-T. and T.S. designed the model, developed the methods and analysed the data. J.A.-L. and A.G.-T. performed the calculations. J.A-L, A.G.-T. and J.I.C. wrote the manuscript with input from all authors. J.I.C. and P.Z. conceived the study and were in charge of the overall direction and planning.

**Competing interests** The authors declare no competing interests.
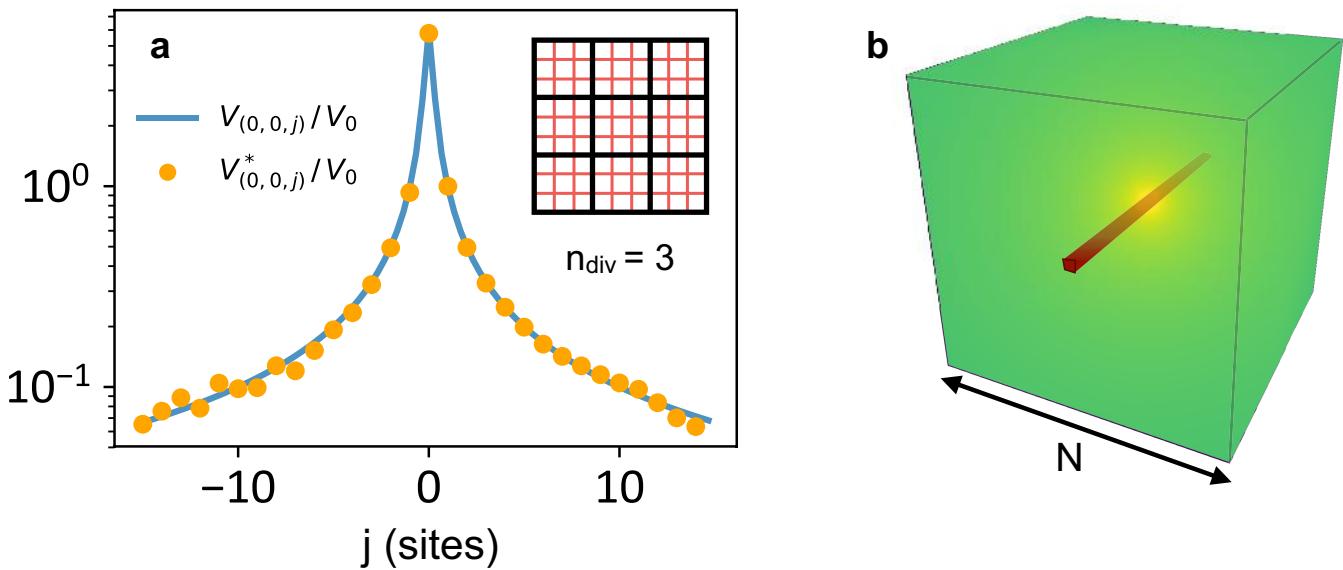
**Extended Data Fig. 1 | Results of the G-S algorithm. a, b,** We apply the G-S algorithm to identify the phase mask associated with a holographic 3D Coulomb potential on a lattice of $N^3$ sites. Fixing the origin at the central site, we choose the nucleus position as $\mathbf{n} = [(2n_{div})^{-1}, 0, 0]$ (the first coordinate is shifted so that the lattice induces a natural cutoff). **a,** An axial central cut of the potential (yellow markers) in direction $z$ (see aligned set of sites in **b**; red), created by a phase mask composed of $(n_{div}N) \times (n_{div}N)$ cells for $n_{div} = 3$ (see inset), compared to the objective Coulomb potential (blue solid line). In step (ii) of the algorithm, the Ewald sphere is discretized using a parallel projection, as in ref. [31]. The field is initiated with random phases. Parameters: $N = 30$ and 7,000 iterations of the G-S algorithm. **b,** Location of the axial cut shown in **a**.
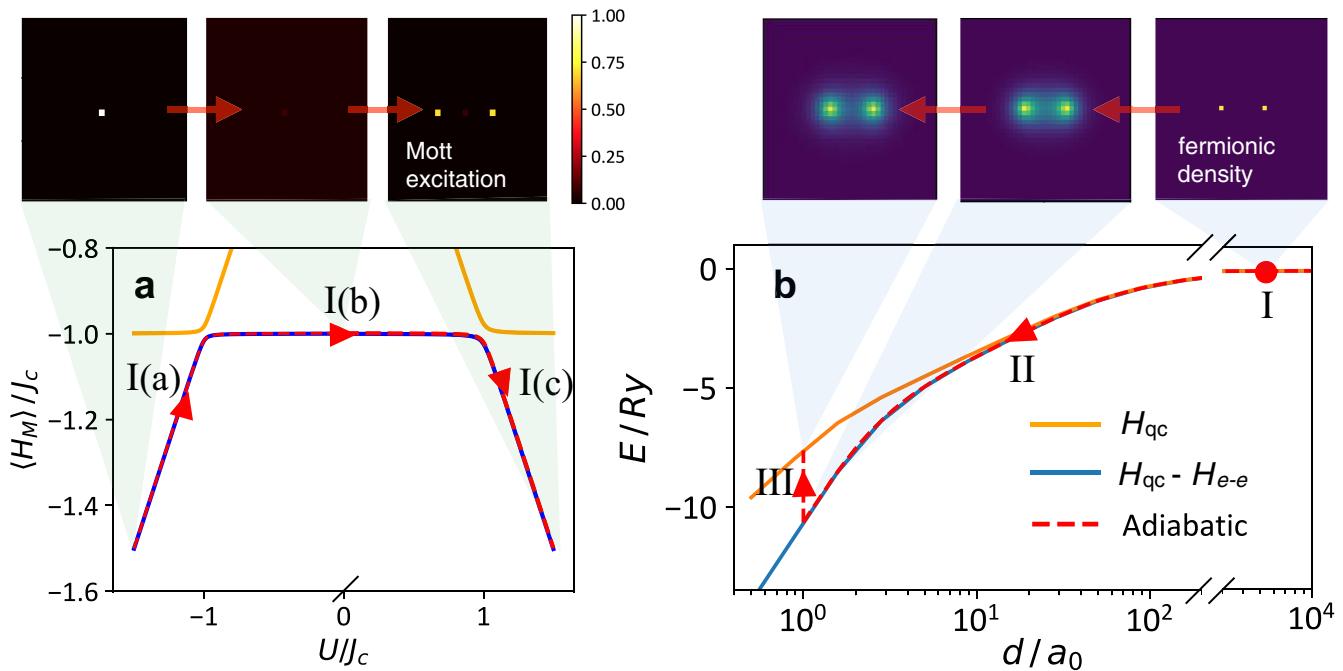
**Extended Data Fig. 2 | Numerical simulation of the adiabatic preparation of the ground state of H₂ with the simulator (particularized for a 2D lattice).** Red dashed lines follow the adiabatic evolution of the initial state $|\psi_0\rangle$ and arrows point to the direction of evolution. **a**, Preparation of the bosonic state through steps I(a)–I(c) (see Methods and Extended Data Table 1). Continuous lines indicate the exact energy of the two lowest energy states. For the adiabatic evolution we use Trotterized time as $\Delta t U = 0.5$ and evolution with $|\Delta U|/(U^2 \Delta t) = 3 \times 10^{-4}$. **b**, Steps II–III of the preparation of the fermionic state. In the simulation, we use

the Trotterized time evolution in intervals of $\Delta t V_0 = 0.05$. In step II, the kinetic term is adiabatically introduced in steps of $\Delta t_F/(V_0^2 \Delta t) = 0.005$. In step III, the electronic repulsion is tuned up as $\Delta V/(V_0^2 \Delta t) = 0.02$. Here yellow (blue) continuous lines follow the exact energy levels of $H_{qc}$, as calculated by imaginary-time evolution with (without) the effect of electronic repulsion. The top insets show the Mott excitation (**a**) fermionic population (**b**) in the lattice at the times indicated in the figure. The final point of the evolution shown in **b** corresponds to $d/a_0 = 1$. Parameters: $N = 60$, $U/J_c = 1.5$, $d/a = 10$.

**Extended Data Table 1 | Evolution of the main parameters of the system during adiabatic preparation with steps I–III presented in Methods**

| Stage | On-site interaction | Attractive potential | Fermionic hopping | Raman coupling |
|-------|--------------------|--------------------|------------------|----------------|
| I(a) | $U \to 0$ | $0$ | $0$ | $0$ |
| I(b) | $0$ | $V_0$ | $0$ | $0$ |
| I(c) | $0 \to U$ | $V_0$ | $0$ | $0$ |
| II | $U$ | $V_0$ | $0 \to t_F$ | $0$ |
| III | $U$ | $V_0$ | $t_F$ | $0 \to g$ |

To simplify the preparation of $|\psi_1\rangle$, step I (illustrated in Extended Data Fig. 2) has been divided into three consecutive substeps: I(a): starting from state $|\psi_0\rangle$, the on-site interaction $U \to 0$ is adiabatically cancelled; this brings the Mott excitation into a delocalized single excitation shared by all the atoms in the insulator, $\sum_j a_j^\dagger |0\rangle / N^{3/2}$. I(b): as discussed in Methods, one can use a dynamic double-well potential to move the fermionic wavepackets in opposite directions. We note that because fermionic and bosonic species are decoupled at this point ($U = 0$), the Mott excitation will remain in the symmetric state reached in I(a). I(c): we adiabatically restore the on-site repulsion $U \to 0$; the bosonic state then evolves to a superposition localized at the new position of the fermionic atoms, $(a_{n_1}^\dagger + a_{n_2}^\dagger)|0\rangle / \sqrt{2}$.

# LETTER

https://doi.org/10.1038/s41586-019-1613-5

# Measuring the Berry phase of graphene from wavefront dislocations in Friedel oscillations

C. Dutreix[1]*, H. González-Herrero[2,3], I. Brihuega[2,3,4], M. I. Katsnelson[5], C. Chapelier[6] & V. T. Renard[6]*

Electronic band structures dictate the mechanical, optical and electrical properties of crystalline solids. Their experimental determination is therefore crucial for technological applications. Although the spectral distribution in energy bands is routinely measured by various techniques[1], it is more difficult to access the topological properties of band structures such as the quantized Berry phase, $\gamma$, which is a gauge-invariant geometrical phase accumulated by the wavefunction along an adiabatic cycle[2]. In graphene, the quantized Berry phase $\gamma = \pi$ accumulated by massless relativistic electrons along cyclotron orbits is evidenced by the anomalous quantum Hall effect[4,5]. It is usually thought that measuring the Berry phase requires the application of external electromagnetic fields to force the charged particles along closed trajectories[3]. Contradicting this belief, here we demonstrate that the Berry phase of graphene can be measured in the absence of any external magnetic field. We observe edge dislocations in oscillations of the charge density $\rho$ (Friedel oscillations) that are formed at hydrogen atoms chemisorbed on graphene. Following Nye and Berry[6] in describing these topological defects as phase

singularities of complex fields, we show that the number of additional wavefronts in the dislocation is a real-space measure of the Berry phase of graphene. Because the electronic dispersion relation can also be determined from Friedel oscillations[7], our study establishes the charge density as a powerful observable with which to determine both the dispersion relation and topological properties of wavefunctions. This could have profound consequences for the study of the band-structure topology of relativistic and gapped phases in solids.

Wave–particle duality manifests as an oscillatory structure in the static response of conduction electrons to impurities: Friedel oscillations[8]. These appear in various contexts and can, for example, alter the conductance of two-dimensional electron gases[9] or mediate long-range interactions between magnetic impurities[10–12]. Because Friedel oscillations intrinsically result from the quantum interference of electronic waves, they necessarily carry information about the crystalline host materials, which impose constraints on the possible wavefunctions. For instance, the wavelength of Friedel oscillations is inversely proportional to the Fermi wavevector $\mathbf{q}_F$ and can be used to recover the energy
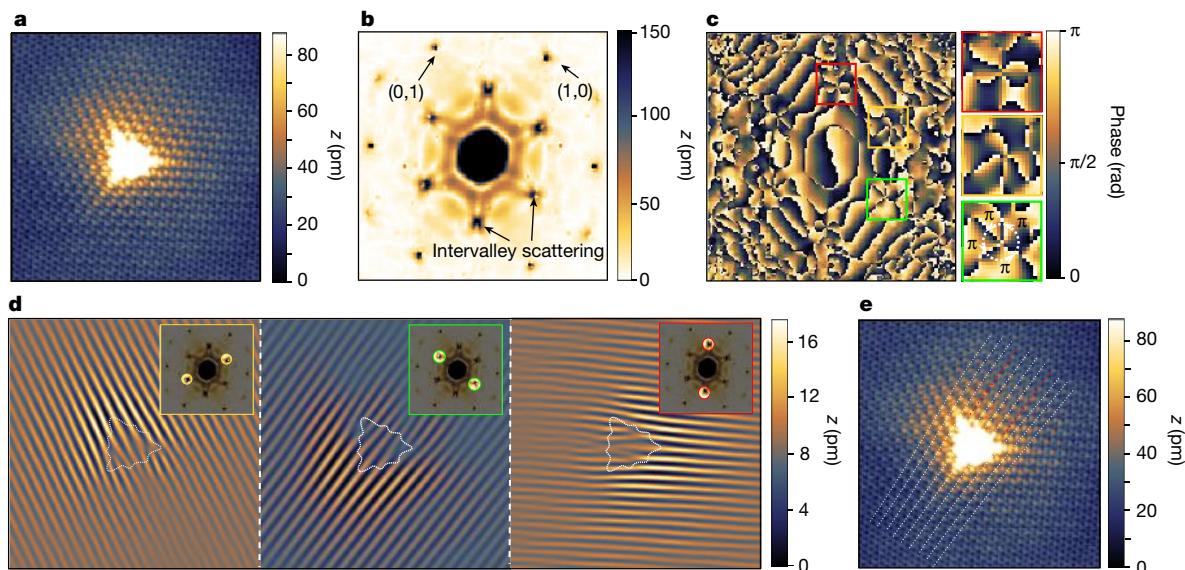


**Fig. 1 | Dislocations in Friedel oscillations near a H atom.**
**a**, Topography STM image of a H adatom at the surface of graphene. The image is 10 nm × 10 nm in size. The tunnelling bias is $V_b = 0.4$ V and the tunnelling current is $i_t = 45.5$ pA. **b**, Modulus of the fast Fourier transform (FFT) of the image in **a**. The points labelled (1, 0) and (0, 1) correspond to the atomic signal. **c**, Phase of the FFT of the image in **a**. Magnifications of the inter-valley backscattering signal are presented on the right, with corresponding border colours red, yellow and green. The phase winds by $4\pi$ around each of these spots; the sharp boundary between bright and dark indicates a phase shift of $\pi$. The FFT images are 62.8 nm$^{-1}$ × 62.8 nm$^{-1}$. **d**, FFT-filtered images of **a** along the three directions of inter-valley scattering. The dotted shape has been added manually to indicate the position of the H atom. The insets show the filters applied in the Fourier space. **e**, Raw image, with dotted lines highlighting the wavefront for one direction of inter-valley scattering. The red dotted lines correspond to the additional wavefronts. Similar results are obtained in the other directions (see Supplementary Information).
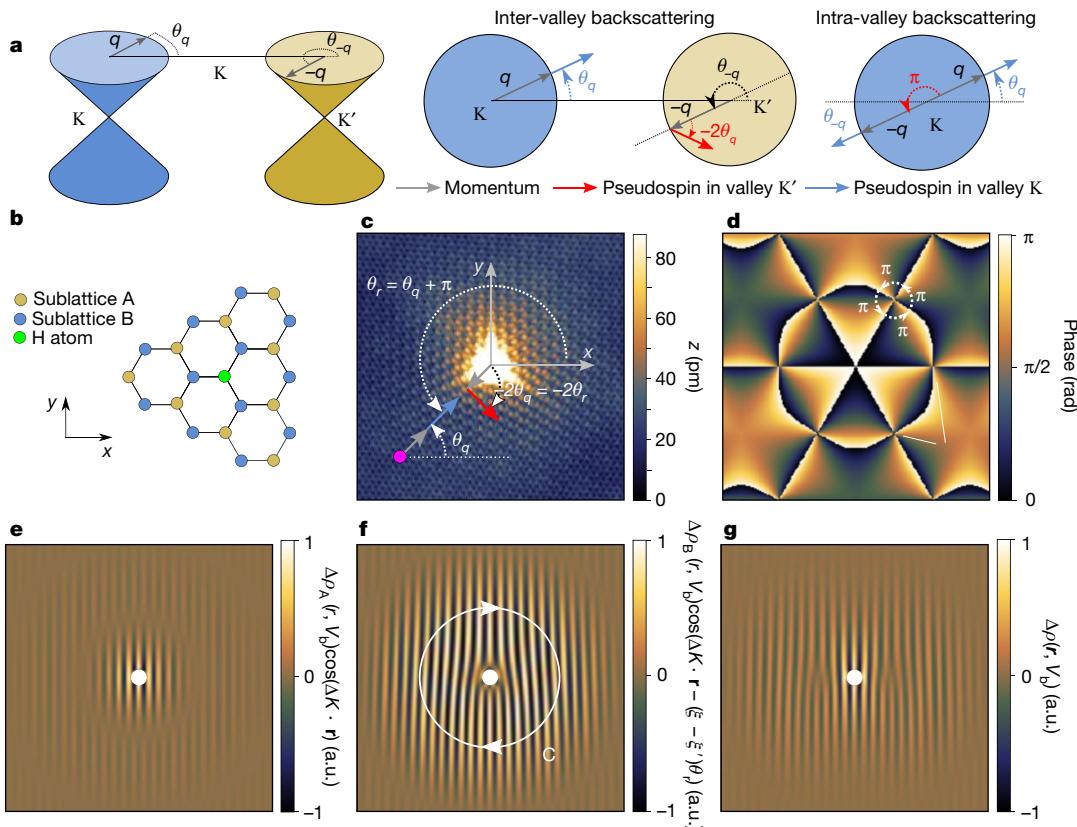
**Fig. 2 | Theoretical description of the dislocations in Friedel oscillations. a**, The backscattering process in graphene. Inter-valley backscattering between wavevector states **q** and −**q** belonging to nearest-neighbour valleys K and K′ leads to a rotation of $-2\theta_{\mathbf{q}}$ of the pseudospin. Intra-valley backscattering rotates the pseudospin by π. **b**, The honeycomb lattice of graphene with a chemisorbed H adatom on sublattice A. **c**, The relation between the STM tip position and the pseudospin rotation in inter-valley backscattering by a H atom. The STM image is the same as in Fig. 1a but is rotated for direct comparison with the theory. The STM tip is represented by the magenta dot. **d**, Phase image of the Fourier transform of the theoretical density modulation $\Delta\rho(\mathbf{r})$ (for comparison with Fig. 1c). The image is 59 nm$^{-1}$ × 59 nm$^{-1}$. **e–g**, The calculated charge density modulation induced by inter-valley scattering on sublattice A (**e**) and on sublattice B (**f**), and the total charge density modulation (**g**). The modulations have been normalized to 1. The images are 10 nm × 10 nm and the theory is integrated from 0 eV to 0.4 eV, as in the STM experiments in Fig. 1. The white disk depicts the H adatom.

dispersion relation from a sequence of energy-resolved scanning tunnelling microscope (STM) images[7]. This has been used to reconstruct the linear dispersion in graphene[13,14]. Friedel oscillations have also been used to demonstrate the existence of the pseudospin of graphene, which arises from the sublattice degree of freedom[14,15]. However, the pseudospin winding, which is directly related to the Berry phase of graphene and characterizes the band structure topology of massless relativistic electrons, has not been retrieved from such STM images.

Figure 1a shows an STM image from our experiments of a H atom chemisorbed on graphene (see Methods and ref. [16] for experimental details). The Fourier transform (Fig. 1b, c) contains signatures of Friedel oscillations associated with the elastic backscattering of massless, relativistic electrons (Dirac electrons) from a given valley K to a nearest-neighbour valley K′ (refs [13–15]). Figure 1d shows the corresponding oscillation in real space after Fourier-filtering the signal for each direction of inter-valley backscattering. Along with the expected inter-valley scattering oscillations, with a wavelength of $\lambda_{\Delta K} = 2\pi/\Delta K \approx 3.7$ Å (where $\Delta K = \mathbf{K}' - \mathbf{K}$ connects two adjacent Dirac points), the filtered images present two dislocations in the vicinity of the H adsorbed atom (adatom). Trained eyes can track them in raw images (Fig. 1e and Supplementary Information). STM imaging after removing the H atoms[16] reveals no structural defects in the graphene, showing that the dislocations appear only in Friedel oscillations. These dislocations allow us to measure the Berry phase, as they are real-space consequences of the pseudospin winding of graphene around the apex of the conical dispersion relation (Dirac cone)—as we will now show.

Friedel oscillations in STM images are dominated by backscattering processes along iso-energy contours[17]. At a given tip position, the amplitude of the Friedel oscillation probed by the STM is governed by the interference of the electronic wave pointing towards the H atom and its reflection from the adatom. As a consequence, the angle $\theta_{\mathbf{q}}$ that parameterizes the momentum **q** of the incident electron is directly related to the angle $\theta_{\mathbf{r}} = \theta_{\mathbf{q}} + \pi$ that indexes the tip orientation at a given position **r** from the H adatom (see Fig. 2a, c in which the angles are defined with respect to the direction $\Delta K$). In graphene, $\theta_{\mathbf{q}}$ also defines the momentum-locked pseudospin of the incident electronic wave in valley K. Intra-valley backscattering involves a rotation of the pseudospin that is always π, so that the interference is destructive at the leading order (Fig. 2a)[15,18,19]. In contrast, backscattering from valley K to valley K′ involves a rotation of the pseudospin by an angle $-2\theta_{\mathbf{q}} = -2\theta_{\mathbf{r}}$ (see Fig. 2a, c), which does not lead to destructive interference but instead leads to the peculiar interference pattern we observe. This pattern is linked to the Berry phase of graphene because circling the STM tip around the impurity is equivalent to circling the momentum **q** of the incident electron on a closed iso-energy contour around the Dirac point in reciprocal space as $\theta_{\mathbf{q}}$ is locked on $\theta_{\mathbf{r}}$ (see Supplementary Video 1). This is analogous to the trajectory of the momentum on a cyclotron orbit, but with the movement of the STM tip replacing the adiabatic transport of electrons in magneto-transport measurements.

More formally, an isolated H adatom constitutes an atomic scatterer that induces both intra- and inter-valley scattering. It may be modelled by an on-site potential $V_0\delta(\mathbf{r})$ (where the amplitude $V_0 \gg 1$ eV and $\delta(\mathbf{r})$ is the Dirac delta function)[20]. Elastic scattering of Dirac electrons on
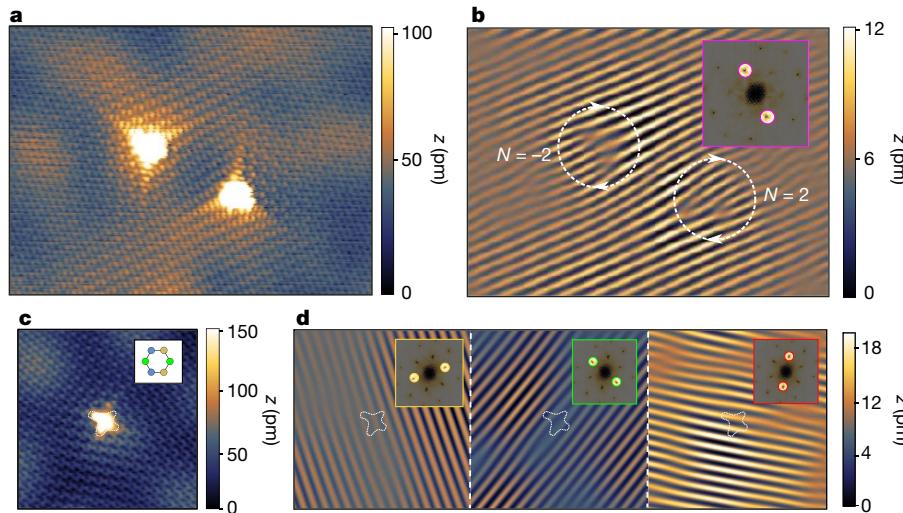
**Fig. 3 | Friedel oscillations around adatoms situated on different sublattices. a**, STM image of the static interferences around two hydrogen adatoms chemisorbed on different sublattices of graphene. The image is 12 nm × 9 nm in size. The tunnelling bias is $V_b = 0.4$ V and the tunnelling current is $i_t = 33.1$ pA. **b**, FFT-filtered image of the image in **a**. The inset shows the filter used. The sign of the topological charge of each phase singularity depends on which sublattice the adatoms belong to. **c**, Topography STM image (6.8 nm × 6.8 nm) of a H-adatom dimer chemisorbed on neighbouring A and B sites (schematic illustration in the inset, same colour code as in Fig. 2b). The tunnelling bias is $V_b = -0.4$ V and the tunnelling current is $i_t = 51.5$ pA. **d**, Image of **c**, FFT-filtered along the three directions of inter-valley scattering. In **b** and **d** the insets are 11 nm$^{-1}$ × 11 nm$^{-1}$.

such a potential has an analytical solution that is non-perturbative in $V_0$ (see Supplementary Information). For an adatom located on sublattice A of the honeycomb lattice of graphene (Fig. 2b) and for a given direction of inter-valley scattering, the elastic scattering yields a modulation of the charge density around the adatom:

$$\Delta\rho(\Delta K, \mathbf{r}, V_b) = \Delta\rho_A(r, V_b)\cos(\Delta K \cdot \mathbf{r})$$
$$+ \xi\xi'\Delta\rho_B(r, V_b)\cos(\Delta K \cdot \mathbf{r} - (\xi - \xi')\theta_\mathbf{r}) \quad (1)$$

Here, the two terms correspond to the charge density modulation on sublattices A and B, respectively, and $V_b$ is the applied STM bias. For intra-valley scattering ($\Delta K = 0$ and the valley index $\xi' = \xi = 1$) the charge modulation is defined entirely by $\Delta\rho_A$ and $\Delta\rho_B$, which describe the universal static response of conduction electrons to impurities—that is, they describe Friedel oscillations[8]. The Friedel oscillations allow the determination of the spectral properties via their $2q_F$-wavevector dependence[13] and have an unconventional decay in graphene because of the π rotation of the pseudospin in intra-valley backscattering[15,18,19]. Their expressions are given in the Supplementary Information. These oscillations have a long period: $\lambda_F/2 = \pi/q_F \approx 5.2$ nm, for $q_F = |\mathbf{q}_F|$ and $\mathbf{q}_F$ fixed by the experimental tunnelling bias $V_b = 0.4$ V.

Extra oscillations appear in equation (1) for inter-valley scattering ($\Delta K \neq 0$, $\xi' = -\xi = -1$). In contrast to usual Friedel oscillations, their wavelength $\lambda_{\Delta K}$ is independent of energy, so that they are not smeared by the integration on the STM bias window (see the experimental proof from d$I$/d$V$ maps in the Supplementary Information). The corresponding modulation of charge density is plotted in Fig. 2e–g for a given direction of inter-valley scattering. Importantly, the angle $-2\theta_\mathbf{r}$, which turns out to be the real-space representation of the pseudospin rotation in inter-valley backscattering (see Supplementary Information), appears as an additional phase shift in the density modulation on sublattice B. It encodes the **q** dependence of the pseudospin and maps its singularity at the Dirac cone apex into a singularity in the real space from which the wavefront dislocation emerges.

The concept of topological defects in waves was introduced by Nye and Berry, who showed that the dislocations in radiowaves echoing off ice sheets in Antarctica resulted from phase singularities in the complex scalar field that describes the wave propagation[6]. Such topological defects in waves are ubiquitous in physics from fluids[21,22] to singular optics[23,24] and condensed matter[25–27]. We follow Nye and Berry in

defining the complex scalar field $\Delta\varrho_B(\mathbf{r}) = |\Delta\varrho_B(\mathbf{r})|e^{i\varphi_B(\mathbf{r})}$, the real part of which describes the Friedel oscillation on sublattice B (the second term in equation (1)). The phase, $\varphi_B(\mathbf{r}) = \Delta K \cdot \mathbf{r} - 2\theta_\mathbf{r}$, is singular at $\mathbf{r} = 0$. It can be regarded as a potential for which the gradient is the sum of a uniform field and a vortex[6]. In our case, the uniform field represents the standing electronic wave associated with inter-valley backscattering, and the vortex represents the perturbation of the wave by the pseudospin rotation. The circulation of this field is the phase accumulated along a closed path C. It is necessarily quantized to a topological number $2\pi N$ (for an integer $N$), because $\Delta\varrho_B(\mathbf{r})$ is a single-valued field and must return to the same observable charge density after circulating along the closed circuit. In singular optics $N$ is called the 'charge' of the phase singularity. It represents the number of additional wavefronts necessary to accommodate for the phase accumulated along C. It is obviously 0 if the closed path does not enclose the phase singularity. For a path enclosing the singularity (Fig. 2f), the gradient circulation of $\varphi_B(\mathbf{r})$ is equal to the winding of $-2\theta_\mathbf{r}$ and hence to that of $-2\theta_\mathbf{q}$. Because the Berry phase in graphene, $\gamma = \pi$, is given by half the winding of $\theta_\mathbf{q}$ (see Supplementary Information) it follows that $2\pi N = 4\gamma$ for a clockwise-oriented contour. The $N = 2$ additional wavefronts seen in Fig. 2f are therefore a signature of the Berry phase in graphene and prove the existence of Dirac cones. We note that, given the quality of the STM image, the winding $4\gamma = 4\pi$ can also be directly retrieved from the phase of the Fourier transform[19] as shown in Figs. 1c and 2d.

The contribution from sublattice A to the total electron density modulation alters only the shape of the dislocation, which is a robust topological feature (see Supplementary Information). The dislocations are shifted from $\mathbf{r} = 0$ in the direction $\Delta K$, in agreement with experiments (Fig. 1d, e and 2g).

The H atom can be placed on a different sublattice, as inferred from the different orientations of the tripod shape[16] of the H signal in Fig. 3a. For a given direction of closed path around the impurity, the sign of $N$ is opposite for the two orientations (Fig. 3b). This is because the two configurations relate to one another via inversion symmetry with respect to the centre of a C–C bond. Because the underlying lattice of graphene is bipartite, this further means that this single-particle topological signature of the sublattice imbalance also relates, via Lieb's theorem[28], to the spontaneous magnetic moments induced by the electron interactions at half filling[16]. In contrast to this many-body effect, the dislocations in Friedel oscillations are independent of doping (see Supplementary

Information). Figure 3c, d shows that if two H atoms are placed on neighbouring carbon atoms there is no dislocation in the inter-valley scattering signal. This results from the annihilation of dislocations of opposite $N$ and illustrates that disorder must break sublattice symmetry (see Supplementary Information).

In quantum mechanics, wavefront dislocations have been predicted for scalar wavefunctions such as the Aharonov–Bohm wavefunction, but were thought to be unobservable owing to the U(1) gauge invariance of the density[21]. We have demonstrated that dislocations appear in the charge density of vectorial wavefunctions, the components of which can interfere with each other by scattering between distant time-reversed valleys. Because wavefront dislocations arise from phase singularities[6] that relate to the topological properties of band structures for vectorial wavefunctions, wavefront dislocations in Friedel oscillations can lead to the identification of relativistic and topological phases, as already established theoretically for rhombohedral graphite[19] and one-dimensional insulators[29]. This method of determining the topological properties of band structures is complementary to transport measurements under a strong magnetic field. However, in contrast to transport measurements, where it destroys quantum Hall measurements, disorder turns out to be an asset here, as long as an area of 100–500 $nm^2$ with a point-like scatterer is available on the surface.

## Online content

1. Sólyom, J. Methods for calculating and measuring the band structure. In *Fundamentals of the Physics of Solids* Vol. 2 (ed. Sólyom J.) 151–194 (Springer, 2009).
2. Berry, M. V. Quantal phase factors accompanying adiabatic changes. *Proc. R. Soc. Lond. A* **392**, 45–57 (1984).
3. Xiao, D. Chang, M.-C., Niu, D. Berry phase effects on electronic properties. *Rev. Mod. Phys.* **82**, 1959–2007 (2010).
4. Novoselov, K. S. et al. Two-dimensional gas of massless Dirac fermions in graphene. *Nature* **438**, 197–200 (2005).
5. Zhang, T., Tan, Y.-W., Stormer, H. L. & Kim, P. Experimental observation of the quantum Hall effect and Berry's phase in graphene. *Nature* **438**, 201–204 (2005).
6. Nye, J. F. & Berry, M. V. Dislocations in wave trains. *Proc. R. Soc. Lond. A* **336**, 165–190 (1974).
7. Crommie, M. F., Lutz, C. P. & Eigler, D. M. Imaging standing waves in a two-dimensional electron gas. *Nature* **363**, 524–527 (1993).
8. Friedel, J. The distribution of electrons round impurities in monovalent metals. *Philos. Mag.* **43**, 153–189 (1952).
9. Zala, G., Narozhny, B. N. & Aleiner, I. L. Interaction corrections at intermediate temperatures: longitudinal conductivity and kinetic equation. *Phys. Rev. B* **64**, 214204 (2001).
10. Ruderman, M. A. & Kittel, C. Indirect exchange coupling of nuclear magnetic moments by conduction electrons. *Phys. Rev.* **96**, 99–102 (1954).
11. Kasuya, T. Theory of metallic ferro- and antiferromagnetism on Zener's model. *Prog. Theor. Phys.* **16**, 45–57 (1956).
12. Yosida, K. Magnetic properties of Cu-Mn alloys. *Phys. Rev.* **106**, 893–898 (1957).
13. Rutter, G. M. et al. Scattering and interference in epitaxial graphene. *Science* **317**, 219–222 (2007).
14. Mallet, P. et al. Role of pseudospin in quasiparticle interferences in epitaxial graphene probed by high-resolution scanning tunneling microscopy. *Phys. Rev. B* **86**, 045444 (2012).
15. Brihuega, I. et al. Quasiparticle chirality in epitaxial graphene probed at the nanometer scale. *Phys. Rev. Lett.* **101**, 206802 (2008).
16. González-Herrero, H. et al. Atomic-scale control of graphene magnetism by using hydrogen atoms. *Science* **352**, 437–441 (2016).
17. Sprunger, P. T., Petersen, L., Plummer, E. W., Lægsgaard, E. & Besenbasher, F. Giant Friedel oscillations on the beryllium(0001) surface. *Science* **275**, 1764–1767 (1997).
18. Cheianov, V. V. & Fal'ko, V. I. Friedel oscillations, impurity scattering, and temperature dependence of resistivity in graphene. *Phys. Rev. Lett.* **97**, 226801 (2006).
19. Dutreix, C. & Katsnelson, M. I. Friedel oscillations at the surfaces of rhombohedral *N*-layer graphene. *Phys. Rev. B* **93**, 035413 (2016).
20. Katsnelson, M. I. *Graphene: Carbon in Two Dimensions* (Cambridge Univ. Press, 2012).
21. Berry, M. V., Chambers, R. G., Large, M. D., Upstill, C. & Walmsley, J. C. Wavefront dislocations in the Aharonov–Bohm effect and its water wave analogue. *Eur. J. Phys.* **1**, 154–162 (1980).
22. Berry, M. V. Making waves in physics. *Nature* **403**, 21 (2000).
23. Dennis, M. R., O'Holleran, K. & Padgett, M. J. Singular optics: optical vortices and polarization singularities. In *Progress in Optics* Vol. 53 (ed. Wolf, E.) 293–363 (Elsevier, 2009).
24. Rafayelyan, M. & Brasselet, E. Bragg–Berry mirrors: reflective broadband q-plates. *Opt. Lett.* **41**, 3972–3975 (2016).
25. Kosterlitz, J. M. & Thouless, D. J. Ordering, metastability and phase transitions in two-dimensional systems. *J. Phys. C* **6**, 1181–1203 (1973).
26. Feynman, R. P. Application of quantum mechanics to liquid helium. In *Progress in Low Temperature Physics* Vol. 1 (ed. Gorter, C. J.) 17–53 (Elsevier, 1955).
27. Abrikosov, A. A. On the magnetic properties of superconductors of the second group. *Sov. Phys. JETP* **5**, 1174–1182 (1957).
28. Lieb, E. H. Two theorems on the Hubbard model. *Phys. Rev. Lett.* **62**, 1201–1204 (1989).
29. Dutreix, C. & Delplace, P. Geometrical phase shift in Friedel oscillations. *Phys. Rev. B* **96**, 195207 (2017).

LETTER RESEARCH

## METHODS

**Sample preparation.** Graphene was grown on silicon carbide (6H-SiC(000$\bar{1}$)) by thermal annealing following the procedure described previously[30]. This leads to the growth of graphene layers, the low-energy physics of which is that of single-layer graphene owing to the decoupling by rotational disorder[31]. The doping of the top graphene layer could be controlled by the number of underlying layers, which is governed by the annealing temperature and time[16]. The results presented in the main text were obtained on a thick multilayer sample (more than five graphene layers) in which the substrate is too far away to dope the layers by charge transfer (see also Supplementary Information). A thinner sample (2–4 graphene layers) was prepared to investigate the effect of doping (see Supplementary Information). Hydrogen atoms were deposited on the surface of the graphene on the SiC substrate by thermal dissociation of $H_2$ in a custom-made H-atom beam source under ultrahigh vacuum conditions[32]. A molecular $H_2$ beam was passed through a hot tungsten filament held at 1,900 K. The pristine graphene substrate was placed 10 cm away from the filament, held at room temperature of around 25 °C during atomic H deposition and subsequently cooled down to 5 K, the temperature at which we carried out all STM and scanning tunnelling spectroscopy experiments presented here. $H_2$ pressure was regulated by a leak valve and fixed to $3 \times 10^{-7}$ torr as measured in the preparation chamber for the present experiments. The atomic H coverage was adjusted by varying the deposition times between 200 s and 60 s, corresponding to final coverages between 0.10 and 0.03 H atoms per $nm^2$.

**STM measurements.** The STM measurements were performed in situ using a custom-made low-temperature STM operating at 5 K under ultrahigh vacuum. Figures 1a and 3a, c were obtained in constant current mode. Conductance spectra and images presented in the Supplementary Information were taken using a lock-in technique, with an a.c. voltage (with a frequency of 830 Hz and amplitude of 1–2 mV root mean square) added to the d.c. sample bias.

## Data availability

The datasets generated and/or analysed during the current study are available from the corresponding author on reasonable request.

30. Varchon, F., Mallet, P., Magaud, L. & Veuillen, J.-Y. Rotational disorder in few-layer graphene films on 6H-SiC(000$\bar{1}$): a scanning tunneling microscopy study. *Phys. Rev. B* **77**, 165415 (2008).
31. Hass, J. et al. Why multilayer graphene on 4H-SiC(000$\bar{1}$) behaves like a single sheet of graphene. *Phys. Rev. Lett.* **62**, 1201–1204 (2008).
32. Hornekær, L. et al. Clustering of chemisorbed H(D) atoms on the graphite (0001) surface due to preferential sticking. *Phys. Rev. Lett.* **97**, 186102 (2006).

# LETTER

# Tuning element distribution, structure and properties by composition in high–entropy alloys

Qingqing Ding[1,9], Yin Zhang[2,9], Xiao Chen[3,9], Xiaoqian Fu[1], Dengke Chen[2], Sijing Chen[1], Lin Gu[4], Fei Wei[3], Hongbin Bei[1], Yanfei Gao[5,6], Minru Wen[2], Jixue Li[1], Ze Zhang[1], Ting Zhu[2]*, Robert O. Ritchie[7,8]* & Qian Yu[1]*

**High-entropy alloys are a class of materials that contain five or more elements in near-equiatomic proportions[1,2]. Their unconventional compositions and chemical structures hold promise for achieving unprecedented combinations of mechanical properties[3–8]. Rational design of such alloys hinges on an understanding of the composition–structure–property relationships in a near-infinite compositional space[9,10]. Here we use atomic-resolution chemical mapping to reveal the element distribution of the widely studied face-centred cubic CrMnFeCoNi Cantor alloy[2] and of a new face-centred cubic alloy, CrFeCoNiPd. In the Cantor alloy, the distribution of the five constituent elements is relatively random and uniform. By contrast, in the CrFeCoNiPd alloy, in which the palladium atoms have a markedly different atomic size and electronegativity from the other elements, the homogeneity decreases considerably; all five elements tend to show greater aggregation, with a wavelength of incipient concentration waves[11,12] as small as 1 to 3 nanometres. The resulting nanoscale alternating tensile and compressive strain fields lead to considerable resistance to dislocation glide. In situ transmission electron microscopy during straining experiments reveals massive dislocation cross-slip from the early stage of plastic deformation, resulting in strong dislocation interactions between multiple slip systems. These deformation mechanisms in the CrFeCoNiPd alloy, which differ markedly from those in the Cantor alloy and other face-centred cubic high-entropy alloys, are promoted by pronounced fluctuations in composition and an increase in stacking-fault energy, leading to higher yield strength without compromising strain hardening and tensile ductility. Mapping atomic-scale element distributions opens opportunities for understanding chemical structures and thus providing a basis for tuning composition and atomic configurations to obtain outstanding mechanical properties.**

In principle, high-entropy alloys (HEAs) should form a single phase with what has been presumed to be a random solid solution[1]. Some HEAs, in particular the CrCoNi-based systems, display exceptional mechanical performance, including high strength, ductility and toughness, particularly at low temperatures[3,5], making them potentially attractive materials for many structural applications. These special characteristics have been attributed to factors that include high entropy, sluggish diffusion and severe lattice distortion[13], issues that are related to the degree of randomness of the solid solution. A fundamental question is whether such solid solutions with multiple principal elements involve unconventional atomic structures or elemental distributions, such as local chemical ordering or clustering, that could affect the defect behaviour and thus enhance mechanical properties. Most theoretical descriptions of solid solutions in HEAs assume that they comprise a random distribution of different atomic species. However, some simulations and more limited experimental results[14–18]

suggest that local variations in chemical composition or even short-range order may exist in HEAs. All five elements in the most studied HEA alloy, CrMnFeCoNi, belong to the first row of transition metals in the Periodic Table, with similar atomic size and electronegativity (a measure of tendency to form intermediate compounds instead of primary solid solutions[19,20]). How would the random alloying effect and associated solid-solution strengthening theories[21,22] be affected if an element from another row were substituted?

To address the above questions, we investigated the atomic-scale element distributions in the CrMnFeCoNi Cantor alloy and a new CrFeCoNiPd HEA (see Methods and Extended Data Fig. 1) by using energy-dispersive X-ray spectroscopy (EDS). To obtain high-resolution EDS maps of individual elements, we used thin and clean samples, a long dwell time and low beam current to reach an appropriately high signal-to-noise ratio (see Methods). Figure 1a presents the atomic-resolution high-angle annular dark field (HAADF) images from transmission electron microscopy (TEM) and corresponding EDS maps of the CrMnFeCoNi alloy with [110] zone axis. On each EDS map for a specific element such as Cr, the brightness of an individual spot increases approximately with the number of Cr atoms in the atomic column along [110] and thus represents the local Cr density. From the EDS maps in Fig. 1a, some random density variations can be seen for all five elements, but Co, Cr and Ni share a more similar degree of homogeneity than Fe and Mn; there is little obvious evidence of assemblies of a particular element. This observation is supported by line profiles of atomic fraction taken from the EDS maps. For example, Fig. 1b shows the line profiles that represent the distribution of individual elements in a $(1\bar{1}1)$ plane projected along the [110] beam direction, which indicate that the atomic fraction of Co, Cr, Ni, Fe or Mn in each projected atomic column randomly fluctuates with small variation. The atomic fraction of Mn has the largest range of variation, and it occasionally reaches a high of about 30% or a low of 12%.

To identify possible repeating patterns obscured by random fluctuations, we calculated the pair correlation functions of the atomic fraction for each element (that is, autocorrelation functions) from the element line profiles represented as the sum of concentration waves[11,12] with a spectrum of wavelengths (see Methods). Figure 1c shows the calculated pair correlation functions of individual elements for concentration wavelength $r$ up to 3.5 nm. Generally, with such correlation function plots, a high peak at wavelength $r$ indicates the incipient concentration wave with characteristic period $r$, and a wide peak reflects the gradual variation of the wave amplitude. However, Fig. 1c shows relatively low and broad peaks in the pair correlation functions for all five elements of the Cantor alloy. This indicates a lack of incipient clustering and thus confirms the observation of random element distributions with small variation from the EDS maps in Fig. 1a.

[1]Center of Electron Microscopy and State Key Laboratory of Silicon Materials, Department of Materials Science and Engineering, Zhejiang University, Hangzhou, China. [2]Woodruff School of Mechanical Engineering, Georgia Institute of Technology, Atlanta, GA, USA. [3]Beijing Key Laboratory of Green Chemical Reaction Engineering and Technology, Department of Chemical Engineering, Tsinghua University, Beijing, China. [4]Beijing National Laboratory for Condensed Matter Physics and Collaborative Innovation Center of Quantum Matter, Institute of Physics, Chinese Academy of Sciences, Beijing, China. [5]Department of Materials Science and Engineering, University of Tennessee, Knoxville, TN, USA. [6]Materials Science and Technology Division, Oak Ridge National Laboratory, Oak Ridge, TN, USA. [7]Materials Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA, USA. [8]Department of Materials Science and Engineering, University of California Berkeley, Berkeley, CA, USA. [9]These authors contributed equally: Qingqing Ding, Yin Zhang, Xiao Chen. *e-mail: ting.zhu@me.gatech.edu; roritchie@lbl.gov; yu_qian@zju.edu.cn
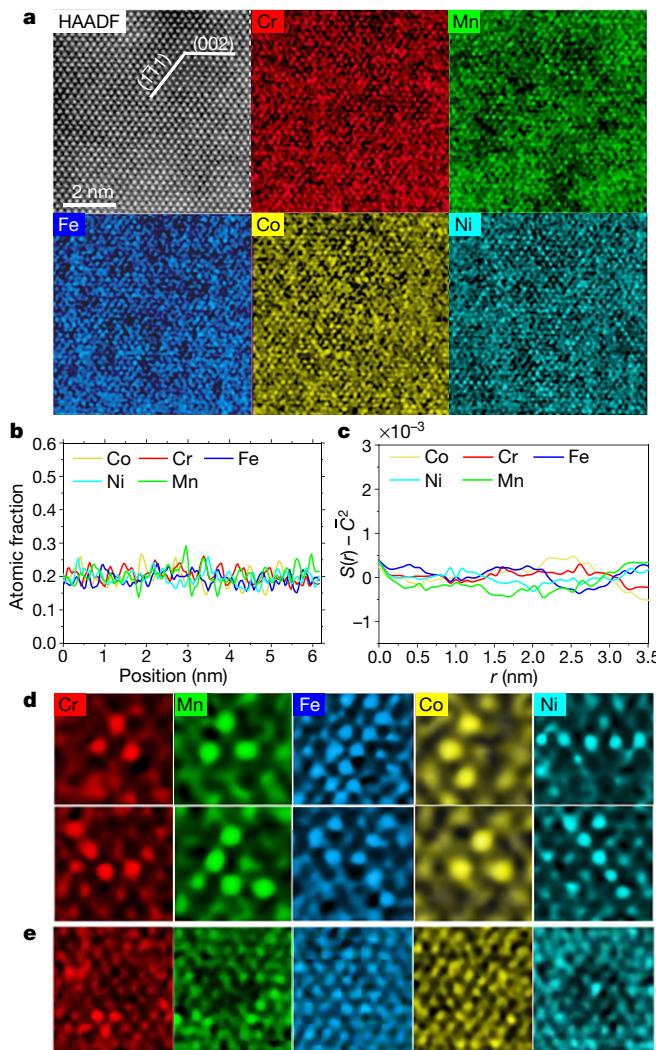
**Fig. 1 | Aberration-corrected TEM imaging and mapping of element distributions in the CrMnFeCoNi Cantor alloy. a**, HAADF image of atomic structure, taken with the [110] zone axis, and corresponding EDS maps for individual elements of Cr, Mn, Fe, Co and Ni. **b**, Line profiles of atomic fraction of individual elements taken from the respective EDS maps in **a**; each line profile represents the distribution of an element in a (1$\bar{1}$1) plane projected along the [110] beam direction. **c**, Plots of pair correlation function $S(r)$ of individual elements against concentration wavelength $r$; $S(r)$ is shifted by $\bar{C}^2$, where $\bar{C}$ denotes the average atomic fraction of the corresponding element. **d**, Magnification of local regions in **a** (all to same scale), showing small groups of neighbouring atomic columns with similar brightness. **e**, Comparison of the local concentration distribution of individual elements for the same region, showing that an Ni-poor region is filled by more Fe and Co than Cr and Mn.

By examining the atomic-resolution EDS maps in detail, certain local groups can be identified in the Cantor alloy, as shown in Fig. 1d. For Cr, Co and Mn, the local group can be as small as $X = 3$, where $X$ is the number of Cr-, Co- or Mn-rich atomic columns forming a triangle or line on either the (1$\bar{1}$1) or (002) planes. Iron tends to aggregate and displays relatively large groups. Nickel is different from the other four elements, as Ni-rich atomic columns prefer to form linear arrays on either (002) or (1$\bar{1}$1) planes. It also appears that the Ni-poor region is filled by more Fe and Co atomic columns rather than Cr and Mn as close neighbours, as illustrated in Fig. 1e. These local groups of atomic columns may be connected to short-range ordering in random solid solutions as suggested by recent modelling studies[18]. However, caution should be used in interpreting these patterns, given a lack of information on the arrangement of elements along each atomic column.
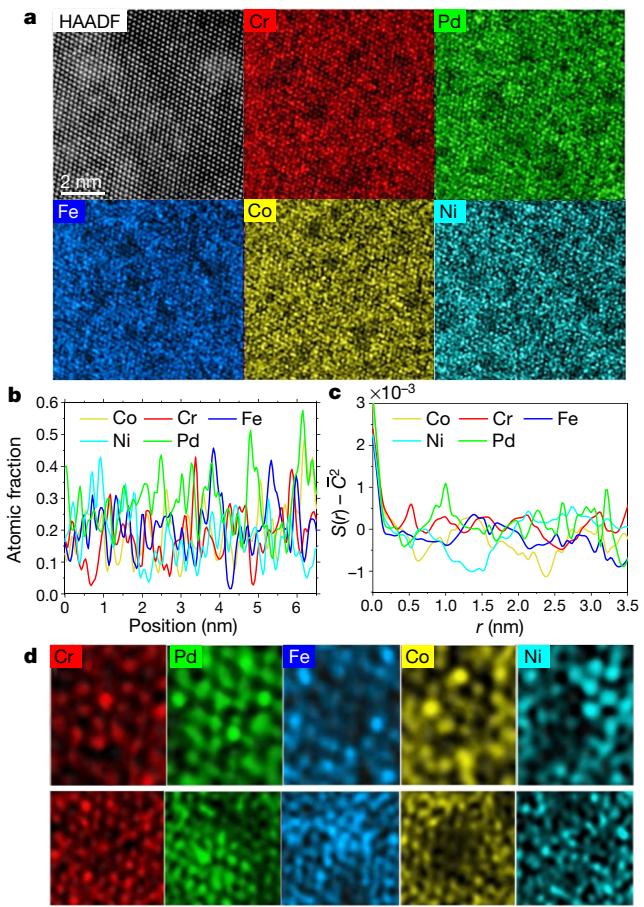


**Fig. 2 | Aberration-corrected TEM imaging and mapping of element distributions in the CrFeCoNiPd alloy. a**, HAADF image of atomic structure, taken with the [110] zone axis, and corresponding EDS maps for individual elements of Cr, Fe, Co, Ni and Pd. **b**, Line profiles of atomic fraction of individual elements taken from respective EDS maps in **a**; each line profile represents the distribution of an element in a (1$\bar{1}$1) plane projected along the [110] beam direction. **c**, Plots of pair correlation functions $S(r)$ of individual elements against concentration wavelength $r$; $S(r)$ is shifted by $\bar{C}^2$, where $\bar{C}$ denotes the average atomic fraction of the corresponding element. **d**, Comparison of the local concentration distribution of individual elements for the same region, showing no obvious preference for specific neighbours.

In contrast to the Cantor alloy, all five elements in the CrFeCoNiPd alloy tend to aggregate. In this HEA, the element distributions in all five EDS maps (Fig. 2a) exhibit strong inhomogeneous fluctuations with local aggregations. The corresponding line profiles (Fig. 2b) show that the atomic fraction of Fe, Co, Ni and Cr can reach 45% but decrease to 2% in regions nearby. For Pd, the fluctuation of its atomic fraction in each atomic column is even larger, with the highest atomic fraction reaching 58%. Despite strong fluctuations in element distributions in Fig. 2a, clusters with well-defined size, spacing and interface cannot be readily identified. Hence, it is more appropriate to characterize these inhomogeneous element distributions as incipient concentration waves[11,12]. The repeating patterns of such waves are obscured by random fluctuations but can be identified by pair correlation analysis. Compared with the earlier results for the Cantor alloy (Fig. 1c), the calculated pair correlation functions for the CrFeCoNiPd alloy (Fig. 2c) reveal a strong correlation peak for Pd, Co and Fe at a concentration wavelength $r$ around 1 nm, a similarly strong correlation peak for Ni at $r$ around 2.5 nm, and a few weak peaks for Cr. As such, these incipient concentration waves repeat their patterns at the length scale larger than that of short-range order, which is usually considered to involve a regular arrangement of atoms within a few atomic neighbour shells, with a repeating length scale less than 0.5 nm (ref. [18]). Moreover, the

correlation function peaks for Pd, Co, Fe and Ni are rather broad, indicating the diffuse interfaces of the corresponding aggregates. Overall, the results from correlation analysis indicate that these inhomogeneous element distributions have salient features of incipient concentration waves[11,12]. In addition, the elements show no obvious preference for specific neighbours (Fig. 2d). Regions that are, for example, Ni-poor, in Fig. 2d, are not filled by a certain element but a mixture of the other four elements.

Comparison of Figs. 1 and 2 indicates that Pd atoms in the CrFeCoNiPd alloy do not simply replace Mn atoms in CrMnFeCoNi but induce substantial changes in the distribution of all elements. In the CrFeCoNiPd alloy, Pd atoms are larger than Fe, Co, Cr and Ni. Moreover, Pd has the largest electronegativity of 6.22 (Mulliken's scale), as opposed to 4.87 for Mn and 4.77 for Cr (the latter being the smallest of the six elements in the two alloys; the electronegativity of other elements is 5.40 for Fe, 5.46 for Co and 5.85 for Ni)[19,20]. For transition metal elements, a larger difference in electronegativity indicates a stronger tendency to form intermediate compounds instead of primary solid solutions. Hence, introducing Pd increases the individual identity of the atoms and promotes aggregations not only of Pd but also of the other four elements, resulting in pronounced chemical inhomogeneities in the alloy. In a coherent structure, an inhomogeneous element distribution inevitably gives rise to non-uniform distribution of lattice strain due to mismatch of atomic sizes. It follows that concentration waves, with characteristic wavelength as small as 1–3 nm in the CrFeCoNiPd alloy, develop through the competing action of lattice strains and concentration gradients on the system energy[11]. Strain-induced composition modulation has been previously observed in binary and ternary alloys[11]. However, the regularity of composition modulation is markedly reduced in the CrFeCoNiPd alloy owing to the increased complexity of local chemical bonding structures among the five constituent elements. Our atomistic Monte Carlo simulations of alloy annealing provide an example of formation of concentration waves across all the constituent elements in a model ternary alloy, which arise because of favoured bonding between certain elements as well as the lattice strain effect (see Methods and Extended Data Fig. 2).

To reveal the impact of the inhomogeneous element distribution on the microscopic deformation mechanisms, we performed in situ TEM straining experiments on the CrFeCoNiPd alloy and compared the results with those of the CrMnFeCoNi alloy. Previous studies on the CrMnFeCoNi alloy showed that in addition to $\frac{1}{2}\langle110\rangle\{111\}$ full dislocations, $\frac{1}{6}\langle112\rangle\{111\}$ partial dislocations were highly active at room temperature[3,4], results that were consistent with the low stacking fault energy $\gamma_{sf}$ of about 30 mJ m$^{-2}$ for the CrMnFeCoNi alloy[23]. In contrast, plastic deformation in the CrFeCoNiPd alloy at room temperature primarily involved $\frac{1}{2}\langle110\rangle\{111\}$ full dislocations. Figure 3a presents an aberration-corrected TEM image of a 60° full dislocation consisting of two partial dislocations. From the measured dislocation core widths, the $\gamma_{sf}$ was estimated to be 66 mJ m$^{-2}$ (Extended Data Figs. 3 and 4), much higher than that of the Cantor alloy[23]. In addition, dislocation motion in CrFeCoNiPd was sluggish, indicative of considerable resistance (that is, high lattice friction) to dislocation glide as indicated in Fig. 3b, which can be related to the pinning effects of pronounced concentration fluctuations in the CrFeCoNiPd alloy.

More importantly, our in situ TEM straining experiments revealed a striking phenomenon of massive cross-slip of screw dislocations in the CrFeCoNiPd alloy from the earliest stages of deformation. Such cross-slip was facilitated by the formation of a sustained dislocation pile-up, as shown in Fig. 3c and Supplementary Video 1. Because dislocations on the primary slip plane experienced high resistance to their motion, a number of dislocations in the pile-up started to cross-slip. Figure 3d shows TEM images of the massive cross-slip that was distributed almost everywhere along the dislocation pile-up. Supplementary Video 2 shows the remarkable 'rainfall'-like process of massive cross-slip. The cross-slipped dislocations frequently underwent secondary cross-slip (Supplementary Video 3), resulting in complex dislocation interactions
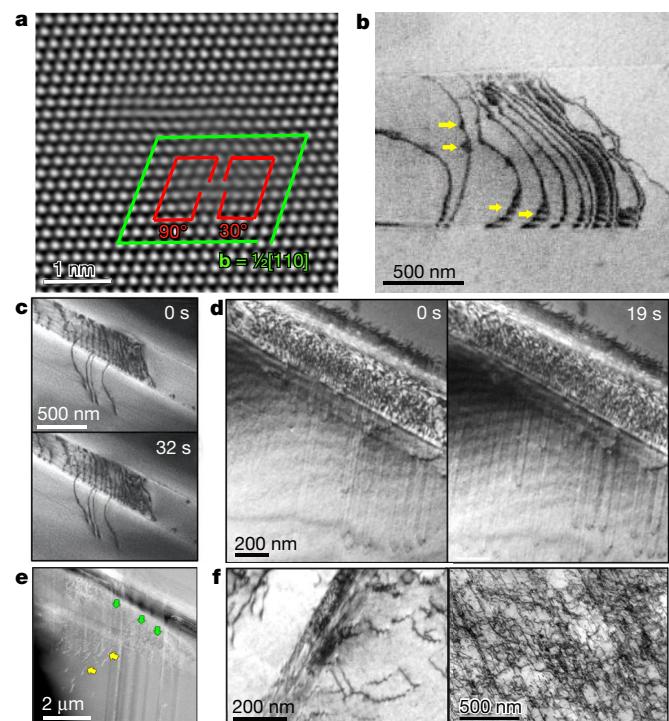
**Fig. 3 | TEM observation of dislocations in the CrFeCoNiPd alloy.**
**a**, HAADF image taken with the $\langle110\rangle$ zone axis, showing the atomic structure of a 60° full dislocation, with the Burgers vector **b** of $\frac{1}{2}[110]$. This 60° dislocation is dissociated into a 30° partial and a 90° partial. The distance between the two partials—that is, the stacking fault width—is as small as about 1 nm. **b**, TEM image taken during in situ TEM straining experiments, showing a dislocation array. Some of the moving dislocation lines exhibit widely separated leading and trailing partials (marked by yellow arrows), showing the temporary pinning of one of the partials. **c**, TEM images showing the sluggish motion of dislocations in a pile-up, where the leading dislocation was obstructed by a strong obstacle. **d**, TEM images at an early time (left image) and a late time (right image) showing massive cross-slip everywhere in the dislocation pile-up. **e**, TEM image showing the activation of new slip systems due to the interaction of intersecting slip bands. Green and yellow arrows respectively indicate the primary and secondary dislocation slip bands. **f**, Post-mortem TEM images showing dislocation microstructures in large-scale samples at the early stage of plastic deformation (left), as well as at the late stage of plastic deformation (right) with an applied large strain of about 30%, where dislocation interactions and multiplication are complex, resulting in a high dislocation density.

(Fig. 3e). Frequent cross-slip and ensuing dislocation interactions promote strain hardening, which is a reliable source of enhanced tensile ductility and toughness[24]. To assess the thin film effect, we performed post-mortem characterization by TEM of the dislocation structures in large-scale samples at different applied strains (Fig. 3f). The results are consistent with in situ TEM observations of frequent cross-slip.

The inhomogeneous element distributions and associated deformation mechanisms can strongly influence the mechanical properties of HEAs. Figure 4a and b shows the measured uniaxial stress–strain curves, at room temperature (293 K) and at liquid nitrogen temperature (77 K), for CrFeCoNiPd and CrMnFeCoNi alloys with two different average grain sizes. At room temperature, the 0.2%-offset yield strength of the CrFeCoNiPd is 410 MPa at a grain size of about 130 μm (and 600 MPa at a grain size of about 5 μm), which is higher than for the CrMnFeCoNi alloy[25] with a similar grain size of 155 μm, and also higher than most reported HEAs with similar grain sizes[3,6,23,25–27] (Fig. 4c and Extended Data Table 1); these strengths are also comparable to those of advanced high-strength steels[28–30]. Furthermore, continuous steady strain hardening is achieved in CrFeCoNiPd, which is much higher than that in other single-phase HEAs with similar grain
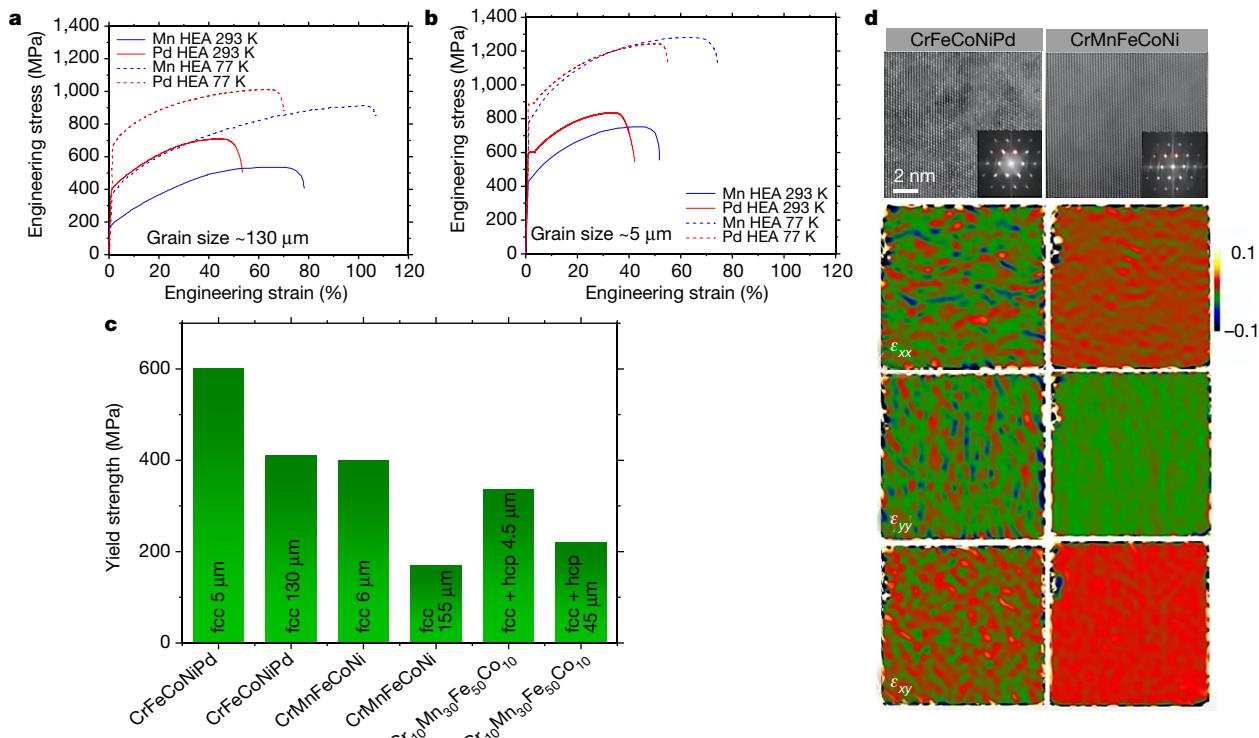
**Fig. 4 | Comparison of mechanical properties of the CrFeCoNiPd alloy with other CrCoNi-based HEAs. a**, Uniaxial tensile stress–strain curves measured at room temperature (293 K) and at liquid nitrogen temperature (77 K) for CrFeCoNiPd (marked as Pd HEA) and CrMnFeCoNi (marked as Mn HEA) with an average grain size of about 130 $\mu$m. **b**, Same as **a** except for an average grain size of about 5 $\mu$m. **c**, Comparison of yield strength between the CrFeCoNiPd alloy and other related HEAs (see the yield strength data in Extended Data Table 1), which have the pure fcc phase or combined fcc and hexagonal close-packed (hcp) phases. **d**, Comparison of atomic strain distribution between the CrFeCoNiPd and CrMnFeCoNi alloys, based on HAADF image and corresponding maps of horizontal normal strain ($\varepsilon_{xx}$), vertical normal strain ($\varepsilon_{yy}$) and shear strain ($\varepsilon_{xy}$).

sizes at ambient temperature[25,31,32]. The elongation to failure is 56% at a grain size of about 130 $\mu$m (and 44% at about 5 $\mu$m grain size). The strain-hardening exponent is similar to that of other CrCoNi-based HEAs, but the hardening of CrFeCoNiPd operates at much higher stresses (Extended Data Fig. 5), leading to an exceptional combination of strength, strain hardening and ductility.

Based on the above experimental results, further insights into the composition–structure–property relationships of the CrCoNi-based HEAs are obtained through comparison of the CrFeCoNiPd and Cantor alloys. Compared with Mn, Pd plays multi-faceted roles in material hardening, including solid-solution hardening due to increased size and modulus mismatches; tuning stacking fault energy; and an increase in obstacle hardening related to aggregations of all five elements. More specifically, according to the Labusch theory[21,22], the degree of solid-solution hardening is largely determined by the size and modulus mismatches between the alloying and matrix atoms. From X-ray diffraction measurements, the lattice mismatch between CrFeCoNiPd (3.67 Å) and CrFeCoNi (3.57 Å) is about 3%, whereas that between CrMnFeCoNi (3.56 Å) and CrFeCoNi is less than 1%. The shear modulus mismatch between CrFeCoNiPd (approximately 89 GPa) and CrFeCoNi (approximately 82 GPa) is also about three times that between CrMnFeCoNi (approximately 80 GPa) and CrFeCoNi. Therefore, based on Labusch's theory, the CrFeCoNiPd alloy should have a stronger effect of solid-solution hardening than CrMnFeCoNi. Second, the Pd atoms in the CrFeCoNiPd alloy lead to more cross-slip by raising the average value of $\gamma_{sf}$ relative to the CrMnFeCoNi alloy. In addition, introducing Pd leads to the aggregation of all five elements. The inhomogeneous element distribution modifies the local value of $\gamma_{sf}$ in the dislocation core, an effect that could lower the effective energy barrier for cross-slip[33]. Third, the inhomogeneous element distribution also modifies the distribution of lattice friction, resulting in stronger resistance to dislocation motion

than in the Cantor alloy. As shown by the strain maps (see Methods) in Fig. 4d, the atomic strain fields in the CrMnFeCoNi alloy are more uniform, whereas substantial atomic strain fluctuations exist in the CrFeCoNiPd alloy. Although the distribution of atomic strain fields appears random at the nanoscale, the tensile and compressive strain fields alternate, which presumably leads to large local internal stresses and thus resistance to dislocation glide.

To further demonstrate the tunability of element distributions in HEAs, we used the element Al to replace Mn and obtained a face-centred cubic (fcc) $Cr_{20}Fe_{20}Co_{18}Ni_{30}Al_{12}$ alloy. Pronounced composition fluctuations in the form of concentration waves and frequent cross-slip were again observed (Extended Data Fig. 6). As a further example of the formation of concentration waves by tuning the alloy composition, we added 5 at% W into a medium-entropy alloy of CrCoNi. The results (Extended Data Fig. 7) show that W atoms can effectively reduce the homogeneity of chemical distribution and enhance the strengthening effect.

The atomic-scale mapping of chemical distribution and associated correlation analysis (including autocorrelation and cross-correlation) open opportunities for resolving the nanoscale chemical structures not only in HEAs but also in other solid solutions more generally. The insight gained into the relationships between chemical structure, microstructure and properties may provide a fundamental basis for tuning compositions and atomic configurations to produce new deformation mechanisms and mechanical properties in HEAs.

## Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-019-1617-1.

1. Yeh, J. W. et al. Nanostructured high-entropy alloys with multiple principal elements: Novel alloy design concepts and outcomes. *Adv. Eng. Mater.* **6**, 299–303 (2004).
2. Cantor, B., Chang, I. T. H., Knight, P. & Vincent, A. J. B. Microstructural development in equiatomic multicomponent alloys. *Mater. Sci. Eng. A* **375–377**, 213–218 (2004).
3. Gludovatz, B. et al. A fracture-resistant high-entropy alloy for cryogenic applications. *Science* **345**, 1153–1158 (2014).
4. Zhang, Z. J. et al. Nanoscale origins of the damage tolerance of the high-entropy alloy CrMnFeCoNi. *Nat. Commun.* **6**, 10143 (2015).
5. Gludovatz, B. et al. Exceptional damage-tolerance of a medium-entropy alloy CrCoNi at cryogenic temperatures. *Nat. Commun.* **7**, 10602 (2016).
6. Li, Z. M., Pradeep, K. G., Deng, Y., Raabe, D. & Tasan, C. C. Metastable high-entropy dual-phase alloys overcome the strength–ductility trade-off. *Nature* **534**, 227–230 (2016).
7. Lei, Z. et al. Enhanced strength and ductility in a high-entropy alloy via ordered oxygen complexes. *Nature* **563**, 546–550 (2018).
8. Yang, T. et al. Multicomponent intermetallic nanoparticles and superb mechanical behaviors of complex alloys. *Science* **362**, 933–937 (2018).
9. Zhang, Y. et al. Microstructures and properties of high-entropy alloys. *Prog. Mater. Sci.* **61**, 1–93 (2014).
10. Miracle, D. B. & Senkov, O. N. A critical review of high entropy alloys and related concepts. *Acta Mater.* **122**, 448–511 (2017).
11. Khachaturian, A. G. *Theory of Structural Transformations in Solids* (Dover, 2013).
12. Gyorffy, B. L. & Stocks, G. M. Concentration waves and Fermi surfaces in random metallic alloys. *Phys. Rev. Lett.* **50**, 374–377 (1983).
13. Gludovatz, B., George, E. P. & Ritchie, R. O. Processing, microstructure and mechanical properties of the CrMnFeCoNi high-entropy alloy. *JOM* **67**, 2262–2270 (2015).
14. Widom, M., Huhn, W. P., Maiti, S. & Steurer, W. Hybrid Monte Carlo/molecular dynamics simulation of a refractory metal high entropy alloy. *Metall. Mater. Trans. A* **45**, 196–200 (2014).
15. Tamm, A., Aabloo, A., Klintenberg, M., Stocks, M. & Caro, A. Atomic-scale properties of Ni-based FCC ternary, and quaternary alloys. *Acta Mater.* **99**, 307–312 (2015).
16. Zhang, F. X. et al. Local structure and short-range order in a NiCoCr solid solution alloy. *Phys. Rev. Lett.* **118**, 205501 (2017).
17. Ma, Y. et al. Chemical short-range orders and the induced structural transition in high-entropy alloys. *Scr. Mater.* **144**, 64–68 (2018).
18. Ding, J., Yu, Q., Asta, M. & Ritchie, R. O. Tunable stacking fault energies by tailoring local chemical order in CrCoNi medium-entropy alloys. *Proc. Natl Acad. Sci. USA* **115**, 8919–8924 (2018).
19. Watson, R. E. & Bennett, L. H. Transition metals: *d*-band hybridization, electronegativities and structural stability of intermetallic compounds. *Phys. Rev. B* **18**, 6439–6449 (1978).
20. Cottrell, A. H. *Concepts of the Electron Theory of Alloys* (IOM Communications, 1998).
21. Labusch, R. A statistical theory of solid solution hardening. *Phys. Status Solidi* **41**, 659–669 (1970).
22. Nabarro, F. R. N. Theory of solution hardening. *Philos. Mag.* **35**, 613–622 (1977).
23. Okamoto, N. L. et al. Size effect, critical resolved shear stress, stacking fault energy, and solid solution strengthening in the CrMnFeCoNi high-entropy alloy. *Sci. Rep.* **6**, 35863 (2016).
24. Hart, E. W. Theory of tensile test. *Acta Metall.* **15**, 351–355 (1967).
25. Otto, F. et al. The influences of temperature and microstructure on the tensile properties of a CoCrFeMnNi high-entropy alloy. *Acta Mater.* **61**, 5743–5755 (2013).
26. Wu, Z., Bei, H., Pharr, G. M. & George, E. P. Temperature dependence of the mechanical properties of equiatomic solid solution alloys with face-centered cubic crystal structures. *Acta Mater.* **81**, 428–441 (2014).
27. Liu, W. H. et al. Ductile CoCrFeNiMo$_x$ high entropy alloys strengthened by hard intermetallic phases. *Acta Mater.* **116**, 332–342 (2016).
28. Kim, S.-H., Kim, H. & Kim, N. J. Brittle intermetallic compound makes ultrastrong low-density steel with large ductility. *Nature* **518**, 77–79 (2015).
29. Jacques, P. J., Furnémont, Q., Lani, F., Pardoen, T. & Delannay, F. Multiscale mechanics of TRIP-assisted multiphase steels. I. Characterization and mechanical testing. *Acta Mater.* **55**, 3681–3693 (2007).
30. Bouaziz, O., Allain, S., Scott, C. P., Cugy, P. & Barbier, D. High manganese austenitic twinning induced plasticity steels: a review of the microstructure properties relationships. *Curr. Opin. Solid State Mater. Sci.* **15**, 141–168 (2011).
31. Ma, S. G. et al. Superior high tensile elongation of a single-crystal CoCrFeNiAl$_{0.3}$ high-entropy alloy by Bridgman solidification. *Intermetallics* **54**, 104–109 (2014).
32. Zaddach, A. J., Scattergood, R. O. & Koch, C. C. Tensile properties of low-stacking fault energy high-entropy alloys. *Mater. Sci. Eng. A* **636**, 373–378 (2015).
33. Nöhring, W. G. & Curtin, W. A. Cross-slip of long dislocations in FCC solid solutions. *Acta Mater.* **158**, 95–117 (2018).

## METHODS

**Experimental.** The CrMnFeCoNi alloy was processed as previously reported[25]. Using the same method, the CrFeCoNiPd alloy was produced by arc-melting pure Fe, Co, Ni, Cr and Pd metals (>99.9% purity). To ensure thorough mixing of all elements, the arc-melted buttons were flipped and re-melted at least five times, followed by drop-casting into Cu moulds to produce rectangular ingots with the dimensions $12.7 \times 12.7 \times 70$ mm$^3$. The rectangular bars were homogenized in vacuo at 1,200 °C for 24 h before rolling into 1.8-mm-thick plates at room temperature. Single fcc phase was obtained and confirmed by X-ray diffraction (Extended Data Fig. 1). Equiaxed grain microstructures with average grain sizes of about 130 µm and 5 µm were obtained by recrystallizing at 1,150 °C for 1 h and 20 min in vacuum, respectively.

Atomic-resolution EDS mapping was performed with an aberration-corrected scanning transmission electron microscope (STEM, FEI Titan Cubed Themis G2 300) operated at 300 kV with a convergence semi-angle of 23.6 mrad. The microscope was equipped with a DCOR plus spherical aberration corrector for the electron probe which was aligned before every experiment by using a gold standard sample. The experiments were performed with the following aberration coefficients: A1 = 1.41 nm; A2 = 11.5 nm; B2 = 22.2 nm; C3 = 2.05 µm; A3 = 525 nm; S3 = 177 nm; A4 = 8.81 µm, D4 = 2.39 µm, B4 = 13.2 µm, C5 = −3.95 mm, A5 = 295 µm, S5 = 111 µm, and R5 = 102 µm, ensuring 0.06 nm resolution under normal conditions. The beam current was set between 25 pA and 30 pA. The dwell time was 1 µs per pixel with a map size of $256 \times 256$ pixels; a complete process of EDS mapping took roughly 1.5 h to reach an appropriately high signal-to-noise ratio. The samples were thinned down by jet polish and Ar ion cleaning. Pre-cracked 3 mm TEM grids were pulled to fracture for making ultra-thin areas.

For in situ TEM straining tests, samples were prepared using a twin-jet electropolisher in an acetic acid solution containing 10 vol% perchloric acid at 30 mA and 10 °C. The electron-transparent area was placed on the rectangular hole of the tension substrate and the sample was glued to the tension substrate. We then mounted the tensile sample onto a Gatan single-tilt straining holder with two screws. In situ straining experiments were conducted in an FEI Tecnai G2 F20 TEM operating at 200 kV. The in situ straining tests were conducted in a displacement-control mode.

For bulk mechanical tests, flat tensile specimens with a dog-bone shape and a gauge length of 10 mm were cut from the recrystallized sheets by electrical discharge machining. The tensile axis was perpendicular to the rolling direction. The gauge sections were carefully ground down to 600-grit SiC paper. Uniaxial tensile tests were performed with a screw-driven tensile testing machine (Instron) at a strain rate of $10^{-3}$ s$^{-1}$ and temperatures of 77 K or 293 K. For the 77 K tests, the specimens and grips were entirely immersed in a bath of liquid nitrogen. Room-temperature tests were performed in a laboratory room environment.

Atomic strain maps were obtained using the geometric phase analysis method[34]. We took over 50 aberration-corrected TEM images of the same region and averaged the corresponding atomic positions, so as to minimize the possible influence of vibration during scanning on the strain maps. Although such atomic strain analysis is affected by camera resolution, the qualitative difference in atomic strain maps between the two alloys studied is substantial, indicating the qualitative difference in lattice distortion between the two alloys.

**Correlation analysis.** To find the repeating patterns of element distribution, we calculated pair correlation functions from line profiles of atomic fraction taken from the EDS maps. For each correlation function calculation, seven line profiles for an individual element were used to obtain the ensemble average. For example, the line profile of the atomic fraction of Cr, $C_{Cr}(x)$, is measured as a function of spatial coordinate $x$ along the [112] direction. The corresponding pair correlation function is defined as $S_{Cr-Cr}(r) = \langle C_{Cr}(x) \cdot C_{Cr}(x+r) \rangle$, where $C_{Cr}(x) \cdot C_{Cr}(x+r)$ is the product between the atomic fractions of Cr at two points with separation $r$, and the symbol $\langle \rangle$ denotes the ensemble average of $C_{Cr}(x) \cdot C_{Cr}(x+r)$ over all possible positions $x$ for a fixed $r$. According to the Wiener–Khintchine theorem[35], the pair correlation function for each element (that is, autocorrelation function) can be related to the power spectral density of the line profile of atomic fraction via the Fourier transform. As such, when the line profiles are considered to consist of a series of concentration waves[12,13] with a spectrum of wavelengths, the high peak at wavelength $r$ on the correlation function plot indicates the incipient concentration wave with the characteristic period $r$. Hence, pair correlation functions can be used to identify the primary repeating patterns on line profiles obscured by random fluctuations.

To calculate the pair correlation function using the discrete data points of a line profile, the above-defined correlation function can be expressed as $S_{Cr-Cr}(r) = \frac{1}{N} \sum_{r < (y-x) < r+\Delta r} C_{Cr}(x) \cdot C_{Cr}(y)$ where $N$ is the total number of pairs of $C_{Cr}(x)$ and $C_{Cr}(y)$; here $C_{Cr}(y)$ denotes the atomic fraction of Cr at the coordinate

$y$ that falls in the range $r < (y-x) < r + \Delta r$, that is, when $y$ is close to $x + r$ but can vary within a small range of $\Delta r$. In our calculations, $\Delta r$ was taken as 0.66 Å. In this way, the correlation function peak associated with the periodic lattice (that is, with the period of the lattice constant) can be filtered out on the pair correlation function plot, revealing the repeating patterns whose period is larger than the lattice constant (that is, beyond the nearest neighbour distance). Owing to concentration fluctuations, there is a certain variability of the average value of atomic fraction for each element. To aid comparison between pair correlation functions of different elements, the plot of $S_{Cr-Cr}(r)$ for each element is shifted by the square of the average atomic fraction, $\bar{C}_{Cr}^2$. The final plot of $S_{Cr-Cr}(r) - \bar{C}_{Cr}^2$ can be used to identify the periodic patterns obscured by random fluctuations.

**Monte Carlo simulation.** We conducted atomistic Monte Carlo simulations of alloy annealing to gain insights into the mechanisms underlying the concentration waves revealed by atomic-resolution EDS mapping. That is, we first set up the fcc structure of a model equiatomic alloy with a random element distribution. The cubic supercell had a side length of 10.7 nm and contained a total of 108,000 atoms. Periodic boundary conditions were applied to the supercell. At a given annealing temperature, this structure was relaxed to lower the system energy by element rearrangement through the Monte Carlo algorithm implemented by the molecular dynamics code LAMMPS[36]. During this simulated annealing process, concentration waves developed across all elements, mainly owing to different bonding energies/preferences between elements. These concentration waves, with a characteristic wavelength of about 2 nm, correspond to a mixture of two coexisting phases with different compositions. The simulated EDS maps and pair correlation function plots strongly resemble our experimental results. Further, Monte Carlo simulations show that the wavelengths of simulated concentration waves vary with annealing temperature, indicating the tunability of chemical structures by annealing temperature.

More specifically, as shown in Extended Data Fig. 2a, we set up an initial random alloy structure for a model ternary alloy, referred to as the ABC alloy. The interatomic potential of this alloy was developed to model the CrCoNi alloy[37] and gives different bonding energies and preferences among the three constituent elements. As a result, this model alloy system favours the formation of two coexisting fcc phases with different compositions: that is, one fcc phase predominantly consists of a mixture of elements A and B, and the other fcc phase is primarily composed of element C. Such two-phase chemical structures arise mainly from the favoured chemical bonding between elements A and B. Starting from a random alloy structure in Extended Data Fig. 2a, the Monte Carlo simulation at an annealing temperature of 800 K resulted in the relaxed atomic structure in Extended Data Fig. 2b. Clearly, a mixture of two coexisting phases has developed, with a complex morphology of aggregates of each phase. Extended Data Fig. 2c shows the simulated EDS map for element C based on the relaxed structure in Extended Data Fig. 2b, for a sample thickness of 8 nm along the [110] zone axis, similar to the TEM sample. It is seen that the concentration of element C varies with a characteristic wavelength of about 2 nm. The other two elements also exhibit similar concentration wave patterns in simulated EDS maps. The characteristic wavelengths around about 2 nm for all the elements are confirmed by their pair correlation function peaks in Extended Data Fig. 2d.

To summarize, our Monte Carlo simulations demonstrate the formation of concentration waves across all the constituent elements in a model alloy. This primarily arises from the favoured bonding between certain transition metal elements that is caused by their large difference in electronegativity. The different atomic sizes help to maintain the stability of these concentration waves after more than $10^6$ Monte Carlo steps of structural relaxation, owing to the competing effects of lattice strains and concentration gradients on the system energy.

**Reporting summary.** Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

### Data availability

All data generated or analysed during this study are included in the published article and Supplementary Information, and are available from the corresponding authors upon reasonable request.

34. Hÿtch, M. J., Snoeck, E. & Kilaas, R. Quantitative measurement of displacement and strain fields from HREM micrographs. *Ultramicroscopy* **74**, 131–146 (1998).
35. Chatfield, C. *The Analysis of Time Series: An Introduction* 6th edn (Chapman and Hall/CRC, 2003).
36. Plimpton, S. Fast parallel algorithms for short-range molecular dynamics. *J. Comput. Phys.* **117**, 1–19 (1995).
37. Li, Q., Sheng, H. & Ma, E. Strengthening in multi-principal element alloys with local-chemical-order roughened dislocation pathways. *Nat. Commun.* **10**, 3563 (2019).

**Competing interests** The authors declare no competing interests.

**Additional information**
**Supplementary information** is available for this paper at https://doi.org/10.1038/s41586-019-1617-1.
**Correspondence and requests for materials** should be addressed to T.Z., R.O.R. or Q.Y.
**Peer review information** *Nature* thanks Elena Pereloma and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.
**Reprints and permissions information** is available at http://www.nature.com/reprints.
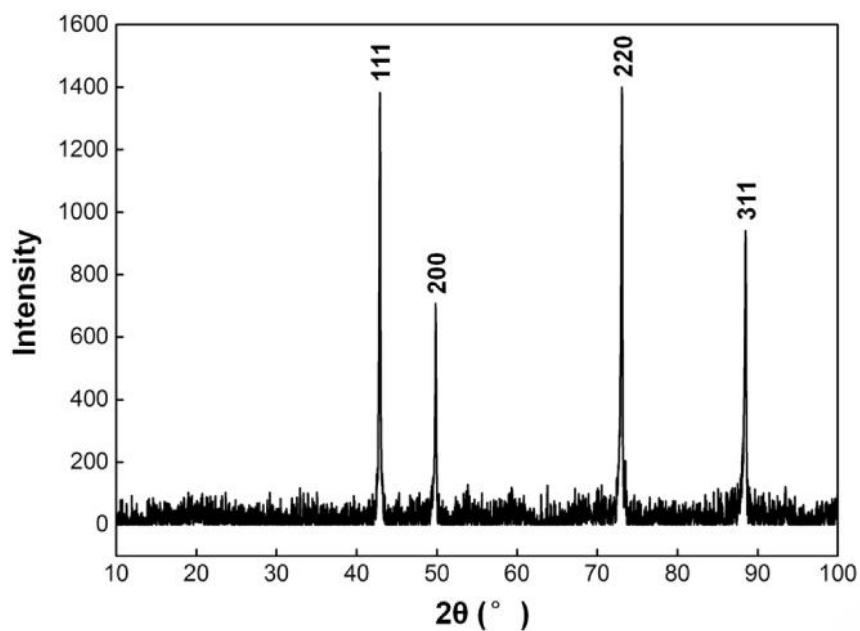
**Extended Data Fig. 1 | X-ray diffraction characterization** showing the single-phase signal of the fcc structure of the CrFeCoNiPd alloy.
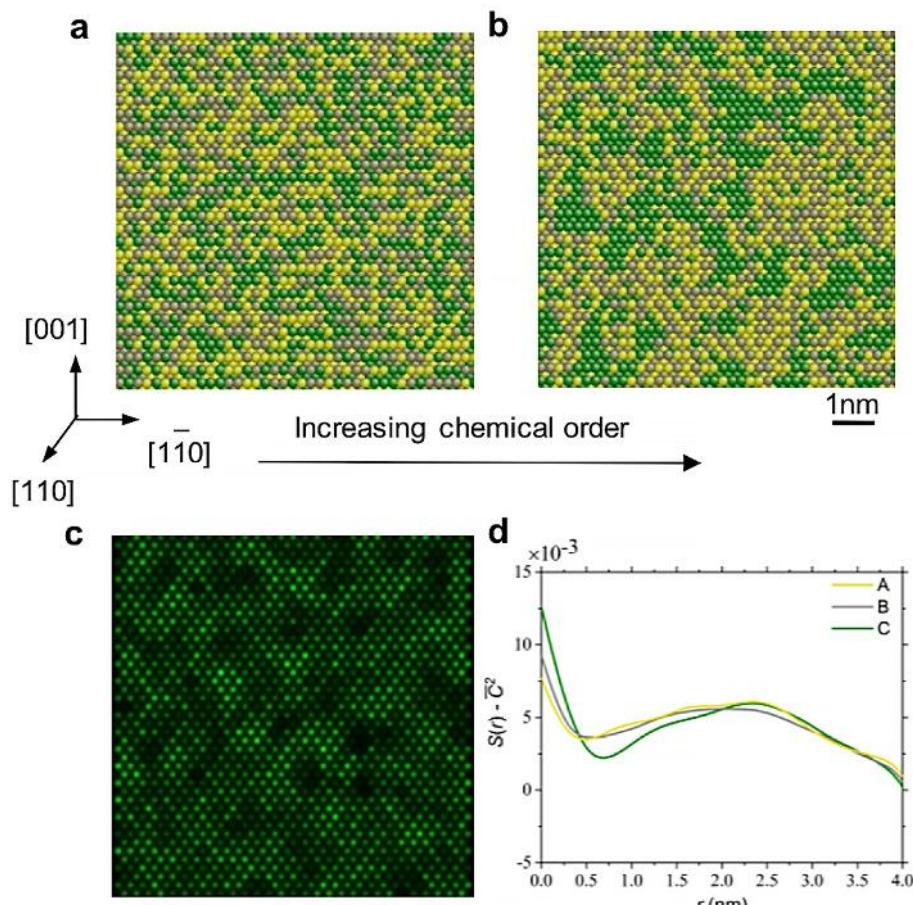
**Extended Data Fig. 2 | Atomistic Monte Carlo simulation.** Simulation shows formation of concentration waves in a model equiatomic ternary alloy under annealing at a temperature of 800 K. **a**, Initial fcc structure with a random distribution of the three constituent elements; yellow, grey and green atoms represent A, B and C elements, respectively. **b**, Relaxed structure showing the formation of a mixture of an element-A/B dominant phase (mixed yellow and grey atoms) and an element-C dominant phase (green atom clusters). **c**, Simulated EDS map for element C based on the structure in **b**. **d**, Plots of pair correlation functions $S(r)$ of individual elements against concentration wavelength $r$.
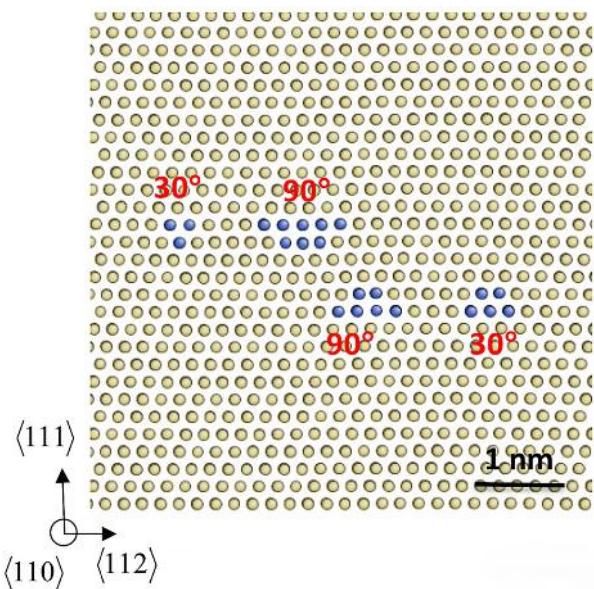
**Extended Data Fig. 3 | Atomic structure of a simulated dislocation dipole.** The structure consists of two closely spaced 60° dislocations of opposite signs in an fcc Ni single crystal, for comparison with similar dislocation core structures in Fig. 3a. Atoms are coloured by their coordination numbers (CN = 12, yellow; CN = 11, blue), so as to display 30° and 90° partial dislocations (atoms in blue) in the core of an extended 60° full dislocation.
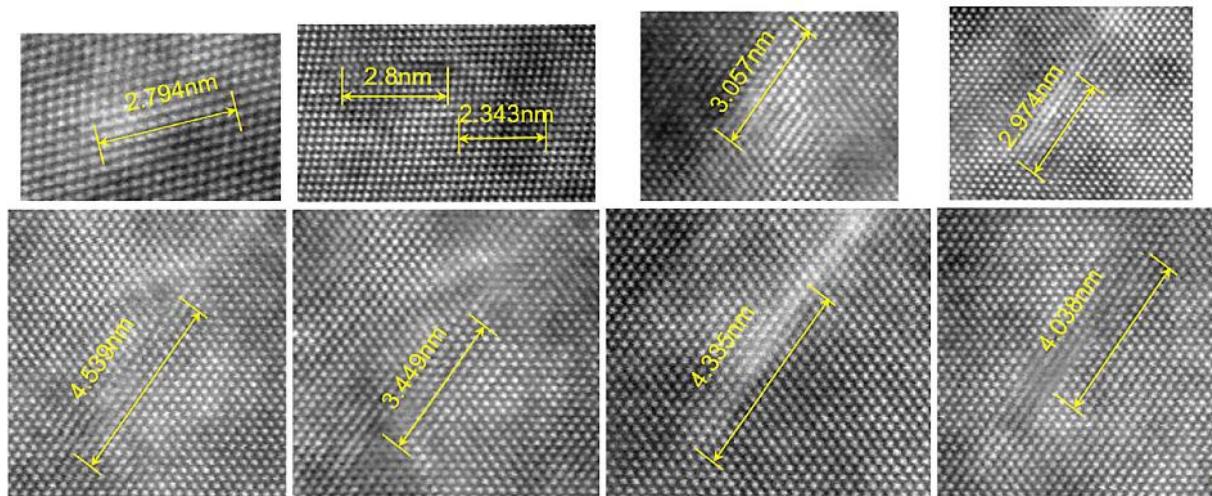
**Extended Data Fig. 4 | High-resolution TEM images of the cores of dissociated 60° dislocations in the CrFeCoNiPd alloy.** The number in each image indicates the measured stacking fault width in the core of dissociated dislocation. The average stacking fault width is $d = 3.37$ nm. The stacking-fault energy $\gamma_{sf}$ can be estimated as

$\gamma_{sf} = \frac{\mu b_p^2}{8\pi d}\left(\frac{2-\nu}{1-\nu}\right)\left(1 - \frac{2\nu\cos2\theta}{2-\nu}\right)$, where $\theta$ is the angle between the dislocation line and the Burgers vector of the full dislocation, $b_p$ is the length of the Burgers vector of the partial dislocation, $\mu$ is the shear modulus and $\nu$ is Poisson's ratio. The $\gamma_{sf}$ of the CrFeCoNiPd alloy is estimated to be 66 mJ m$^{-2}$.
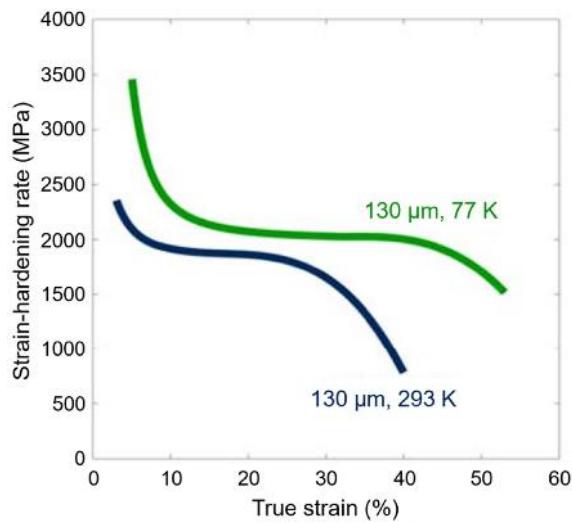
**Extended Data Fig. 5 | Kocks–Mecking plots.** The plots of strain hardening rate against true strain at 293 K and 77 K show the strong hardening capability of the CrFeCoNiPd alloy.
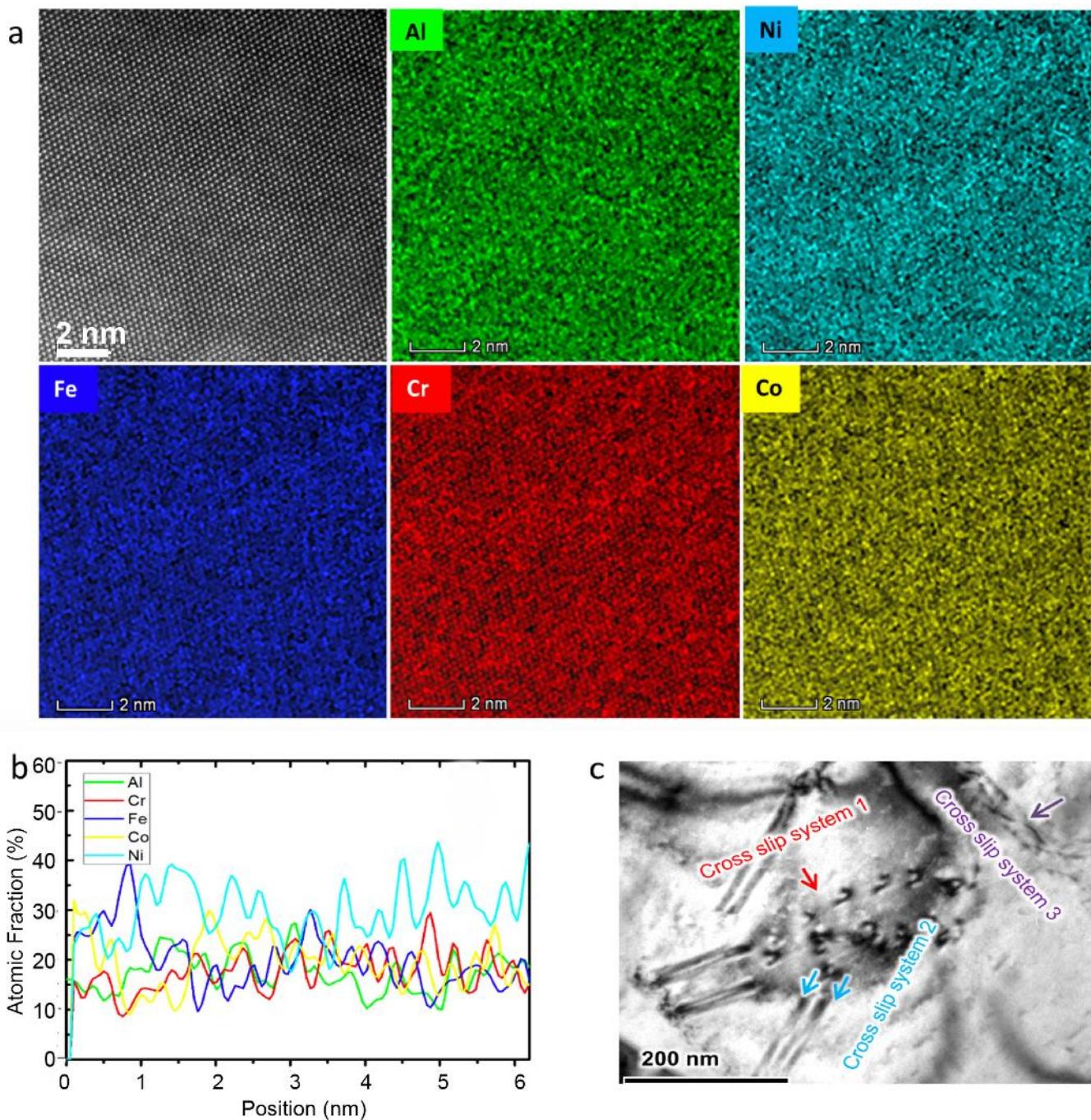
**Extended Data Fig. 6 | Aberration-corrected TEM imaging and mapping of element distributions in the $Cr_{20}Fe_{20}Co_{18}Ni_{30}Al_{12}$ alloy. a**, HAADF images and associated EDS maps (taken along the [110] zone axis) for individual elements of Cr, Fe, Co, Ni and Al. **b**, Line profiles of atomic fraction of individual elements taken from respective EDS maps in **a**; each line profile represents the distribution of an element in a (002) plane projected along the [110] beam direction. **c**, Cross-slip of dislocations in the $Cr_{20}Fe_{20}Co_{18}Ni_{30}Al_{12}$ alloy, from in situ straining experiment.
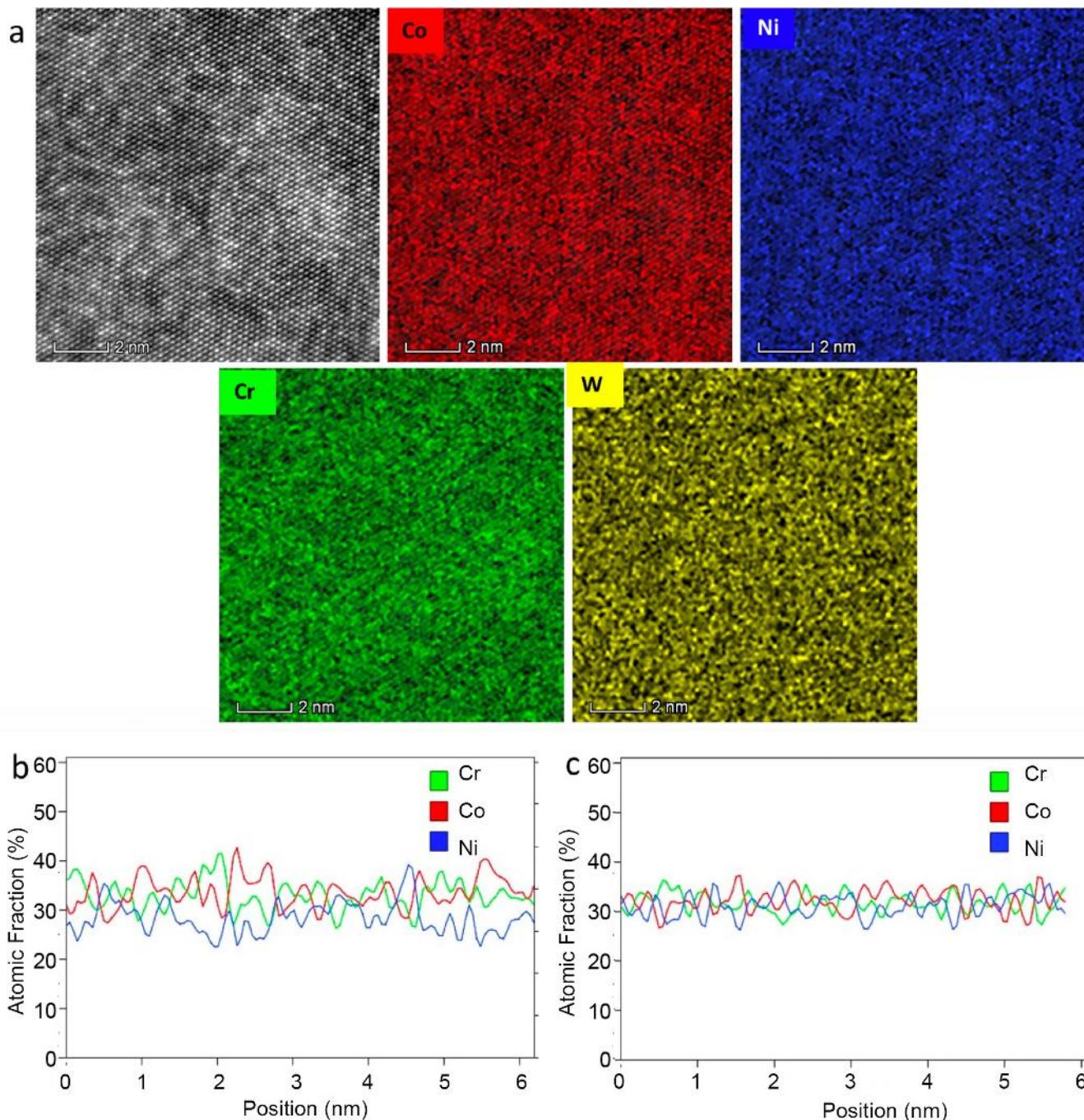
**Extended Data Fig. 7 | Comparison of element distributions in CrCoNi alloy and in CrCoNi alloy containing 5 at% W. a**, HAADF image and corresponding EDS maps of the CrCoNi alloy containing 5 at% W, taken along the [110] zone axis, showing the distribution of individual elements of Cr, Co, Ni and W. **b**, Line profiles of atomic fraction of elements Cr, Co and Ni taken from respective EDS maps in **a** for the CrCoNi alloy containing 5 at% W; each line profile represents the distribution of an element in a $(1\bar{1}1)$ plane projected along the [110] beam direction. **c**, Line profiles of atomic fraction of individual elements taken from the corresponding EDS maps of the CrCoNi alloy; each line profile represents the distribution of an element in a $(1\bar{1}1)$ plane projected along the [110] beam direction.

**Extended Data Table 1 | Properties of the CrMnFeCoNi, CrFeCoNiPd, $Cr_{10}Mn_{30}Fe_{50}Co_{10}$ and CrCoNi alloys**

| Alloy | Grain size (μm) | Temperature (K) | $\sigma_y$ (MPa) | $\sigma_{UTS}$ (MPa) | $\varepsilon_f$ (%) | $n$ |
|---|---|---|---|---|---|---|
| CrMnFeCoNi [3] | 6 | 293 | 410 | 763 | 57 | 0.41 |
| CrMnFeCoNi [3] | 6 | 77 | 759 | 1280 | 71 | 0.36 |
| CrFeCoNiPd | 135 | 293 | 410 | 710 | 56 | 0.33 |
| CrFeCoNiPd | 135 | 77 | 666 | 1012 | 74 | 0.39 |
| CrFeCoNiPd | 5 | 293 | 604 | 835 | 44 | 0.38 |
| CrFeCoNiPd | 5 | 77 | 900 | 1244 | 60 | 0.38 |
| $Cr_{10}Mn_{30}Fe_{50}Co_{10}$ [6] | 4.5 | 293 | 337 | 872 | 74 | n/a |
| $Cr_{10}Mn_{30}Fe_{50}Co_{10}$ [6] | 45 | 293 | 220 | 730 | 51 | n/a |
| CrCoNi [5] | 5-50 | 293 | 440 | 884 | 73 | 0.40 |
| CrCoNi [5] | 5-50 | 77 | 657 | 1311 | 90 | 0.40 |

Yield strength ($\sigma_y$), ultimate tensile strength ($\sigma_{UTS}$), elongation to failure strain ($\varepsilon_f$) and strain-hardening exponent ($n$) of the CrMnFeCoNi, CrFeCoNiPd, $Cr_{10}Mn_{30}Fe_{50}Co_{10}$ and CrCoNi alloys at room and cryogenic temperatures.

# nature research

Corresponding author(s): Robert O. Ritchie

Last updated by author(s): Jun 25, 2019

# Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

## Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

| n/a | Confirmed | |
|---|---|---|
| ☒ | ☐ | The exact sample size ($n$) for each experimental group/condition, given as a discrete number and unit of measurement |
| ☒ | ☐ | A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly |
| ☒ | ☐ | The statistical test(s) used AND whether they are one- or two-sided<br>*Only common tests should be described solely by name; describe more complex techniques in the Methods section.* |
| ☒ | ☐ | A description of all covariates tested |
| ☒ | ☐ | A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons |
| ☒ | ☐ | A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals) |
| ☒ | ☐ | For null hypothesis testing, the test statistic (e.g. $F$, $t$, $r$) with confidence intervals, effect sizes, degrees of freedom and $P$ value noted<br>*Give P values as exact values whenever suitable.* |
| ☒ | ☐ | For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings |
| ☒ | ☐ | For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes |
| ☒ | ☐ | Estimates of effect sizes (e.g. Cohen's $d$, Pearson's $r$), indicating how they were calculated |

*Our web collection on statistics for biologists contains articles on many of the points above.*

## Software and code

Policy information about availability of computer code

| | |
|---|---|
| Data collection | n/a |
| Data analysis | n/a |

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

## Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Data associated with the transmission electron microscopy are available from Prof. Qian Yu at Zhejiang University (email: yu_qian@zju.edu.cn).

# Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☐ Life sciences    ☐ Behavioural & social sciences    ☒ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

# Ecological, evolutionary & environmental sciences study design

All studies must disclose on these points even when the disclosure is negative.

| | |
|---|---|
| Study description | The existence of local chemical ordering on the mechanical properties of Cr-CoNi-based high-entropy alloys. |
| Research sample | n/a |
| Sampling strategy | n/a |
| Data collection | Data collection based on information primarily obtained from transmission electron microscopy. |
| Timing and spatial scale | 1/1/2018 - 2/15/2019 |
| Data exclusions | none |
| Reproducibility | All experiments were repeated at least two times. |
| Randomization | n/a |
| Blinding | n/a |

Did the study involve field work? ☐ Yes ☐ No

## Field work, collection and transport

| | |
|---|---|
| Field conditions | n/a |
| Location | n/a |
| Access and import/export | n/a |
| Disturbance | n/a |

# Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

## Materials & experimental systems

| n/a | Involved in the study |
|---|---|
| ☐ | ☐ Antibodies |
| ☐ | ☐ Eukaryotic cell lines |
| ☐ | ☐ Palaeontology |
| ☐ | ☐ Animals and other organisms |
| ☐ | ☐ Human research participants |
| ☐ | ☐ Clinical data |

## Methods

| n/a | Involved in the study |
|---|---|
| ☐ | ☐ ChIP-seq |
| ☐ | ☐ Flow cytometry |
| ☐ | ☐ MRI-based neuroimaging |

## Antibodies

| | |
|---|---|
| Antibodies used | *Describe all antibodies used in the study; as applicable, provide supplier name, catalog number, clone name, and lot number.* |
| Validation | *Describe the validation of each primary antibody for the species and application, noting any validation statements on the manufacturer's website, relevant citations, antibody profiles in online databases, or data provided in the manuscript.* |

## Eukaryotic cell lines

Policy information about cell lines

| | |
|---|---|
| Cell line source(s) | *State the source of each cell line used.* |

| Authentication | *Describe the authentication procedures for each cell line used OR declare that none of the cell lines used were authenticated.* |
|---|---|
| Mycoplasma contamination | *Confirm that all cell lines tested negative for mycoplasma contamination OR describe the results of the testing for mycoplasma contamination OR declare that the cell lines were not tested for mycoplasma contamination.* |
| Commonly misidentified lines (See ICLAC register) | *Name any commonly misidentified cell lines used in the study and provide a rationale for their use.* |

## Palaeontology

| Specimen provenance | *Provide provenance information for specimens and describe permits that were obtained for the work (including the name of the issuing authority, the date of issue, and any identifying information).* |
|---|---|
| Specimen deposition | *Indicate where the specimens have been deposited to permit free access by other researchers.* |
| Dating methods | *If new dates are provided, describe how they were obtained (e.g. collection, storage, sample pretreatment and measurement), where they were obtained (i.e. lab name), the calibration program and the protocol for quality assurance OR state that no new dates are provided.* |

☐ Tick this box to confirm that the raw and calibrated dates are available in the paper or in Supplementary Information.

## Animals and other organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research

| Laboratory animals | n/a |
|---|---|
| Wild animals | n/a |
| Field-collected samples | n/a |
| Ethics oversight | n/a |

Note that full information on the approval of the study protocol must also be provided in the manuscript.

## Human research participants

Policy information about studies involving human research participants

| Population characteristics | n/a |
|---|---|
| Recruitment | n/a |
| Ethics oversight | n/a |

Note that full information on the approval of the study protocol must also be provided in the manuscript.

## Clinical data

Policy information about clinical studies

All manuscripts should comply with the ICMJE guidelines for publication of clinical research and a completed CONSORT checklist must be included with all submissions.

| Clinical trial registration | n/a |
|---|---|
| Study protocol | n/a |
| Data collection | n/a |
| Outcomes | n/a |

## ChIP-seq

### Data deposition

☐ Confirm that both raw and final processed data have been deposited in a public database such as GEO.

☐ Confirm that you have deposited or provided access to graph files (e.g. BED files) for the called peaks.

| Data access links *May remain private before publication.* | n/a |
|---|---|

| Files in database submission | n/a |
| Genome browser session (e.g. UCSC) | n/a |

## Methodology

| Replicates | n/a |
| Sequencing depth | n/aa |
| Antibodies | n/a |
| Peak calling parameters | n/a |
| Data quality | n/a |
| Software | n/a |

# Flow Cytometry

## Plots

Confirm that:

☐ The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).

☐ The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).

☐ All plots are contour plots with outliers or pseudocolor plots.

☐ A numerical value for number of cells or percentage (with statistics) is provided.

## Methodology

| Sample preparation | n/a |
| Instrument | n/a |
| Software | none |
| Cell population abundance | n/a |
| Gating strategy | n/a |

☐ Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.

# Magnetic resonance imaging

## Experimental design

| Design type | n/a |
| Design specifications | n/a |
| Behavioral performance measures | n/a |

## Acquisition

| Imaging type(s) | n/a |
| Field strength | n/a |
| Sequence & imaging parameters | n/a |
| Area of acquisition | n/a |

| Diffusion MRI | ☐ Used | ☒ Not used |

## Preprocessing

| | |
|---|---|
| Preprocessing software | n/a |
| Normalization | n/a |
| Normalization template | n/a |
| Noise and artifact removal | n/a |
| Volume censoring | n/a |

## Statistical modeling & inference

| | |
|---|---|
| Model type and settings | n/a |
| Effect(s) tested | n/a |

Specify type of analysis: ☐ Whole brain ☐ ROI-based ☐ Both

| | |
|---|---|
| Statistic type for inference<br>(See Eklund et al. 2016) | n/a |
| Correction | n/a |

## Models & analysis

| n/a | Involved in the study |
|---|---|
| ☒ | ☐ Functional and/or effective connectivity |
| ☒ | ☐ Graph analysis |
| ☒ | ☐ Multivariate modeling or predictive analysis |

# LETTER

# Self-coalescing flows in microfluidics for pulse-shaped delivery of reagents

Onur Gökçe[1,5], Samuel Castonguay[2], Yuksel Temiz[1], Thomas Gervais[2,3,4]* & Emmanuel Delamarche[1]*

**Microfluidic systems can deliver portable point-of-care diagnostics without the need for external equipment or specialist operators, by integrating all reagents and manipulations required for a particular assay in one device[1]. A key approach is to deposit picogram quantities of dried reagents in microchannels with micrometre precision using specialized inkjet plotters[2–5]. This means that reagents can be stored for long periods of time and reconstituted spontaneously when adding a liquid sample. But it is challenging to carry out complex operations using multiple reagents, because shear flow enhances their dispersion and they tend to accumulate at moving liquid fronts, resulting in poor spatiotemporal control over the concentration profile of the reconstituted reagents[6]. One solution is to limit the rate of release of reagents into the liquid[7–10]. However, this requires the fine-tuning of different reagents, conditions and targeted operations, and cannot readily produce the complex, time-dependent multireagent concentration pulses required for sophisticated on-chip assays. Here we report and characterize a capillary flow phenomenon that we term self-coalescence, which is seen when a confined liquid with a stretched air–liquid interface is forced to 'zip' back onto itself in a microfluidic channel, thereby allowing reagent reconstitution with minimal dispersion. We provide a comprehensive framework that captures the physical underpinning of this effect. We also fabricate scalable, compact and passive microfluidic structures—'self-coalescence modules', or SCMs—that exploit and control this phenomenon in order to dissolve dried reagent deposits in aqueous solutions with precise spatiotemporal control. We show that SCMs can reconstitute multiple reagents so that they either undergo local reactions or are sequentially delivered in a flow of liquid. SCMs are easily fabricated in different materials, readily configured to enable different reagent manipulations, and readily combined with other microfluidic technologies, so should prove useful for assays, diagnostics, high-throughput screening and other technologies requiring efficient preparation and manipulation of small volumes of complex solutions.**

The key to translating self-coalescence into a useful microfluidic technology is first to elicit and characterize its physical underpinnings (see also Supplementary Information, section 1), in order to demonstrate how air–liquid interfaces can be manipulated such that the characteristic time and length over which dispersion occurs scale with the width, $W$, of the microfluidic channel rather than its length, $L$ (assuming that $W$ is much less than $L$). To implement the concept, we use a shallow channel geometry (whose height, $H$, is much less than $W$) to confine the liquid within a Hele-Shaw cell—a quasi-two-dimensional planar flow model whereby flow is always considered to be locally parabolic and propagating in the direction of the pressure gradient[11]. Within the flow plane, the fluid is further confined laterally by a capillary pinning line (CPL)[12] that acts as a Laplace pressure barrier[13] and geometrically forces the liquid to self-coalesce—that is, to fold onto itself and release its surface free energy to spill over the CPL. Making the CPL straight creates a translational symmetry

in the self-coalescence process that results in the fluid propagating at the velocity of the capillary front. In this process, following the path of least resistance, most of the flow occurs right behind the moving liquid front, where the fluid rapidly comes to a rest (Fig. 1a, b and Supplementary Video 1).

Free-boundary flows are known to be challenging to model and to compute numerically, let alone analytically[14]. A complete solution to the problem can be reached by numerical modelling. Alternatively, recognizing the conformal invariance of the advection–diffusion transport equation in shallow microfluidic devices[15], and that all boundary conditions become regular in the reference frame of the moving meniscus (Fig. 1c), a complete solution can be obtained using a Schwarz–Christoffel mapping[16] (see Supplementary Information, section 1.2). Two simple asymptotic solutions can be further derived via this conformal mapping approach to obtain the main velocity field along the CPL both near its contact point (where $x$ is much less than $W$) and far downstream of the meniscus contact point (where $x$ is much greater than $W$):

$$\overline{V}_{\text{near}}(x,t) \approx U_m + i(U + U_m)(x + U_m t)/r \tag{1}$$

$$\overline{V}_{\text{far}}(x,t) = A(W, \lambda, \theta_w)U_m e^{-\frac{\pi}{W}(x + U_m t + i\lambda W)} \tag{2}$$

where the variables $x$ and $t$ are respectively the horizontal distance from the meniscus tip and the time elapsed; $r = \lambda W/(1 + \cos\theta_w)$ is the radius of the circular meniscus; $U_m$ is the meniscus velocity; $U$ is the flow velocity at the inlet; $A(W, \lambda, \theta_w)$ is a constant geometric factor that depends on the channel width and wall contact angle; $\lambda$ is the width fraction of the channel cross-section at which self-coalescence occurs; $i$ is the imaginary unit; and e is Euler's number (Supplementary Information, section 1.5).

The exponential form of Equation (2) has been extensively studied in the case of Saffman–Taylor viscous fingering in a long straight channel, which yields a similar exponential decay of the velocity field with characteristic decay length $L_d = W/\pi$ and time $t_d = L_d/U_m$ (Fig. 1d)[17]. In the context of self-coalescence, this decay constant explains the short flow lifetime before stagnation is achieved ($L_d$ is roughly 160 μm; $t_d$ is approximately 240 ms; assuming that $W = 500$ μm and $U_m = 1$ mm s$^{-1}$). The analytical results for these two asymptotic behaviours, as well as the full Schwarz–Christoffel mapping solution (see Supplementary Information, section 1.3), reveal a strong match with finite-element-method simulations and experiments, further verifying the model (Fig. 1e). To achieve self-coalescence the CPL can be made straight, but in general can be of any shape, from sharp turns to spirals and slow meanders. This process thus also explains the filling dynamics within capillary pumps[18] and around microfluidic phaseguides[19]—capillary structures often used to control wetting in microsystems. Even more importantly, the process reveals how the phenomena can be precisely triggered and controlled to engineer minimally dispersive flows.

[1]IBM Research–Zurich, Rüschlikon, Switzerland. [2]Department of Engineering Physics, École Polytechnique de Montréal (EPM), Montréal, Québec, Canada. [3]Institut du Cancer de Montréal, Montréal, Québec, Canada. [4]Centre de Recherche du Centre Hospitalier de l'Université de Montréal (CRCHUM), Montréal, Québec, Canada. [5]Present address: Institute of Neuroinformatics, University of Zurich/ETH Zurich, Zurich, Switzerland. *e-mail: thomas.gervais@polymtl.ca; emd@zurich.ibm.com
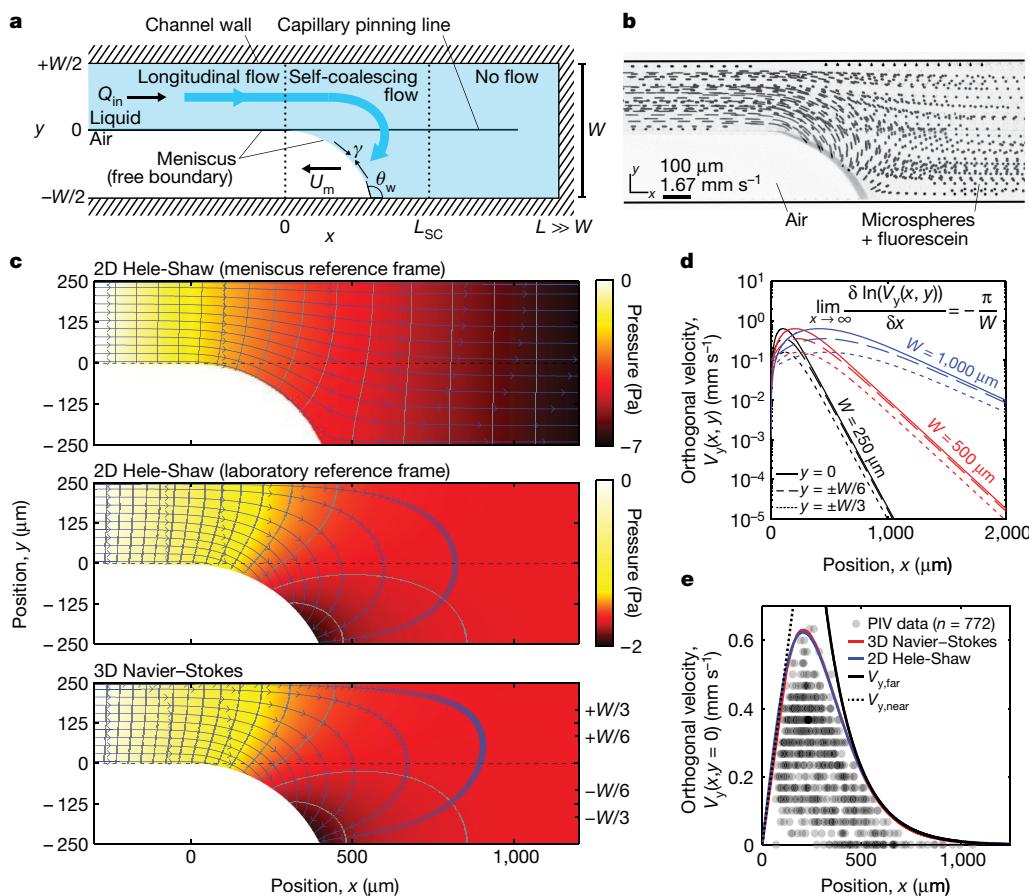
**Fig. 1 | Description and modelling of self-coalescence in a microchannel. a,** Top view of a microchannel (of width $W$ and length $L$) that is partially filled with a liquid. During Stokes flow (at flow rate $Q_{in}$), this setting fosters three flow modes: longitudinal flow (where $x < 0$), self-coalescing flow ($0 < x < L_{SC}$), and stagnancy ('no flow', where $x > L_{SC}$), with $L_{SC}$ being the penetration length of a self-coalescing flow. $U_m$ is the meniscus velocity, $\gamma$ the surface tension, and $\theta_w$ the contact angle. **b,** Image of fluorescent microspheres (diameter 4.8 µm) in 10 µg ml$^{-1}$ fluorescein solution during self-coalescence ($W$ is 500 µm; microchannel height is 50 µm; $Q_{in}$ is 500 nl min$^{-1}$; Re is roughly 0.001; and $\theta_w$ = 116°). Streaks and dots respectively reveal moving and stationary particles. The CPL, implemented with a 5-µm-wide trench, is not visible after background subtraction. **c,** Two-dimensional (2D) Hele-Shaw model (top and middle) and 3D Navier–Stokes simulation (bottom) of self-coalescence in the same channel geometry as in panel **b** and at half the channel height. The colour map shows pressure, and the middle and bottom panels share the same colour map. Blue, velocity streamlines, and grey, isobars. The gaps between arrowheads show the displacement of a particle at regular time intervals (180 ms). **d,** Log plots of the orthogonal velocity ($V_y$) of self-coalescing flows for three channel widths and at different $y$ positions as indicated, highlighting the asymptotic form of the orthogonal velocity component: near the meniscus the flow is transitional, increasing steadily until it reaches a peak and then decaying exponentially downstream of the offset. **e,** Plotted are $V_y$ values for particles from all $z$-planes at $y = 0$ from velocimetry data (semi-transparent dots, $n = 772$ from 12 locations in 3 experiments), along with the maximum $V_y$ predicted by the 2D Hele-Shaw model (blue), the 2D asymptotic model (black, $V_{y,far} = 5.26e^{-\pi x/500}$ mm s$^{-1}$; dotted, $V_{y,near} = 0.0045x$ mm s$^{-1}$) and the 3D Navier–Stokes simulation (red). PIV, particle image velocimetry.

We use the practical implementation of self-coalescence here to control reagent reconstitution under the form of an SCM. The SCM is a microchannel having two CPLs and a vent (Fig. 2a). The first CPL (the leading barrier) ensures longitudinal flow, whereas the second (the diversion barrier) acts as a capillary burst valve[20] and prevents the liquid from exiting the SCM before the SCM is completely filled. A gap between the leading and diversion barriers forces self-coalescence to begin next to the diversion barrier. The effectiveness of reagent reconstitution using self-coalescence becomes apparent when it is compared with reconstitution in a microchannel that lacks a leading barrier, illustrated in Fig. 2b, c using amaranth dye and water. In this case, the longitudinal flow leads to the strong accumulation of amaranth dye near the liquid filling front, which creates a concentration spike that is an order of magnitude higher than the uniform concentration attained when using self-coalescence (Fig. 2d and Extended Data Fig. 1). When self-coalescence is used to reconstitute reagents printed at any fixed point in that region, the Péclet number, Pé, associated with the flows depends on the local velocity above a reagent spot, which peaks linearly (Equation (1)) before decaying exponentially with time as the meniscus

moves away (Equation (2)). Thus Pé $\equiv |V_{far}|H/D$, which approximates to $e^{-\pi Ut/\lambda W}$, making Taylor–Aris dispersion, whose magnitude scales roughly as Pé$^2$, vanishingly small almost immediately after reagents have been wetted, and practically eliminating any form of convective reagent dispersion during reconstitution (Fig. 2b and Supplementary Video 2).

There are numerous ways to implement SCMs in different materials—using, for example, silicon substrates and microfabrication or polymeric materials via hot embossing or injection moulding, and using different geometries such as depressed (trench-like) or protruding (rail-like) CPLs (Extended Data Fig. 2). The material compatibility is further supported by the fact that the contact angle has only a mild influence on self-coalescence, and no influence at all on the exponential-decay length scale (Supplementary Information section 1.5). Capillary pinning can also be achieved by patterning hydrophobic layers[21].

When performing chemistry at the microscale, it is a common requirement to use solutions that have a well defined composition and volumes that range from nanolitres up to a few microlitres. The volume of the SCM determines the maximum volume of the solution
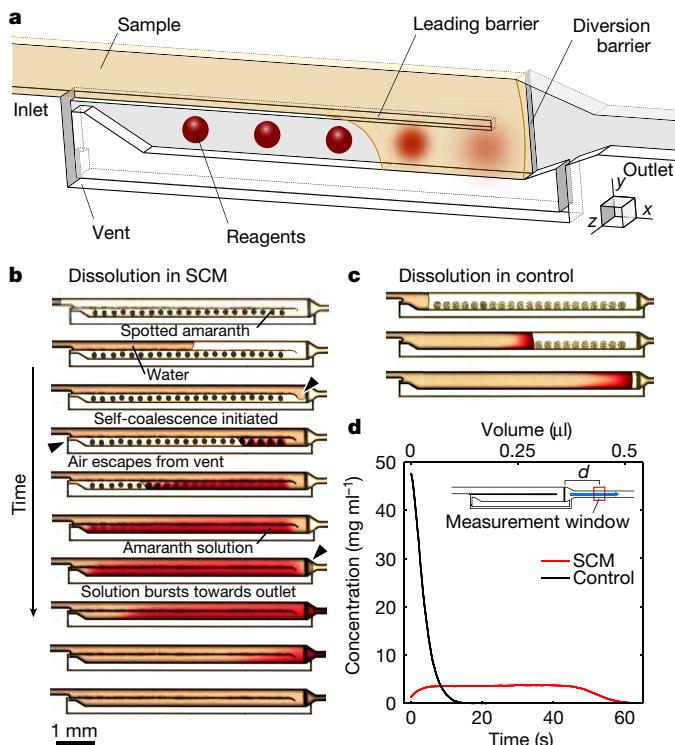
**Fig. 2 | SCM for reagent reconstitution. a**, Illustration of the components forming an SCM and the locations of dried, spotted reagents. **b**, Time series of bright-field microscope images showing the reconstitution of amaranth in an SCM filled with water over the course of 75 s. **c**, Reconstitution of amaranth in a control experiment (with no self-coalescence) shows strong reagent accumulation. **d**, Mean concentration profiles of amaranth solutions using a 375-nl SCM (red, $n = 6$) and a 375-nl control microchannel (black, $n = 6$, $Q_{in} = 500$ nl min$^{-1}$). Inset, the amaranth concentration was measured at $d = 3$ mm downstream of the SCM outlet.



**Fig. 3 | Scalability of SCMs. a**, The width ($W$) and depth ($H_{SCM}$) of an SCM and the length of the deposited reagent pattern ($L_{reagent}$) determine the volume in which a reagent reconstitutes ($V_{reagent}$). The final concentration of the reagent after reconstitution can simply be adjusted with the amount of deposited reagent ($m_{reagent}$). **b**, Mean concentration profiles ($n = 6$; measurement distance $d = 3$ mm) of amaranth, reconstituted in SCMs with varied lateral dimensions ($H_{SCM} = 50$ μm for all; $Q_{in} = 500$ nl min$^{-1}$).

that can be prepared, and the amount of deposited reagent determines its resulting concentration, as demonstrated here with amaranth dye reconstitution in SCMs of different dimensions (Fig. 3 and Extended Data Fig. 3). We tested the stability of leading barriers using different channel geometries and filling conditions (Extended Data Fig. 4). Modelling and experiments reveal simple rules of thumb for designing stable SCMs: wide SCMs are preferred over long SCMs that accumulate more hydraulic resistance during filling.

The preparation of solutions with uniform concentration profiles is perhaps the greatest practical challenge solved by SCMs. In fact, any arbitrary concentration profile can be generated using one or several reagents and specific spotting patterns. However, a broadening of the initial reagent concentration profile is inevitable owing to Taylor–Aris dispersion once the solution sets in motion and exits the SCM (Supplementary Information, section 2.2). The magnitude of the dispersion and the resulting concentration profile at time $t$ and distance $d$ from the diversion barrier can be computed exactly by convolving the spotted reagent profile with the Green's functions that describe reagent profile evolution during SCM filling and reagent delivery (Supplementary Information, section 2.3). The approach yields a powerful design tool, which we use for reagent pulse shaping under Taylor–Aris dispersion in a way that is conceptually analogous to optical amplitude pulse shaping in dispersive media[22], where the spatial frequency distribution of the input concentration signal replaces the frequency content of a light pulse, concentration replaces the pulse amplitude, and the dispersive channel acts as the dispersive optical medium (Supplementary Information, section 2.4).

This modelling strategy for concentration pulse shaping allows us to define spotting patterns that will give specific concentration profiles of

reconstituted amaranth and brilliant blue dye without having to optimize the patterns empirically (Fig. 4). We can readily release reagents consecutively with minimal intermixing (Fig. 4a), achieve steady concentration ranges (Fig. 4b), and merge reagents at preprogrammed distances downstream (Fig. 4c). It is also possible to integrate two different reagents and release them downstream either sequentially with minimum dispersion (Supplementary Information, section 2.5) and a separated profile (Fig. 4d), or mixed with an intercalated profile (Fig. 4e), or with a gradient/counter-gradient profile (Fig. 4f), in close agreement with the theoretical predictions of our concentration pulse shaping model (Extended Data Fig. 5 and Supplementary Information, section 2.6).

SCM variants that have a modified vent to purge a surplus of liquid and no diversion barrier can maintain the distribution profiles of reconstituted reagents for long durations, with minimal diffusion effects. This allows several biochemical reactions to be run in spatially segregated regions inside single SCMs, akin to using individual wells in a microtitre plate to perform simultaneous experiments (Fig. 5). We illustrate this with a multistep enzymatic assay that uses a fluorometric readout to quantify the activity of glucose-6-phosphate dehydrogenase (G6PDH)[23], with a glucose-6-phosphate (G6P) substrate, the co-factors nicotinamide adenine dinucleotide phosphate (NADP$^+$) and magnesium (Mg$^{2+}$), and a fluorescent reporter system that is based on the reduction of resazurin by diaphorase spotted in a first SCM (SCM 1) (Fig. 5a; note that diaphorase is spotted separately from its substrates in order to avoid undesired enzymatic activity during storage of the SCM). When the carrier fluid is introduced, these reagents reconstitute homogeneously and proceed to the next SCM, where reaction kinetics (which depend on the G6PDH concentration) can be characterized using the fluorescence signal of resorufin. SCMs are also powerful tools for calibrating such an enzymatic reaction, where the ambient temperature, slight variations in the amount of deposited reagents, or the possible decay of reagents over time might alter the reaction kinetics. Such a calibration can be performed by spotting a gradient of G6PDH in SCM 2 and measuring the resulting kinetics (Fig. 5a). The amount of deposited G6PDH and the maximum speed of the enzymatic reaction (Fig. 5b) exhibit the expected linear relation, as shown by a calibration curve (Fig. 5c, black). These calibration data match quantification data (Fig. 5c, orange), which were obtained by loading a buffer spiked with various amounts of G6PDH into SCMs containing all other reagents. It was possible to measure in only 2 min as little as 0.75 μU μl$^{-1}$ of
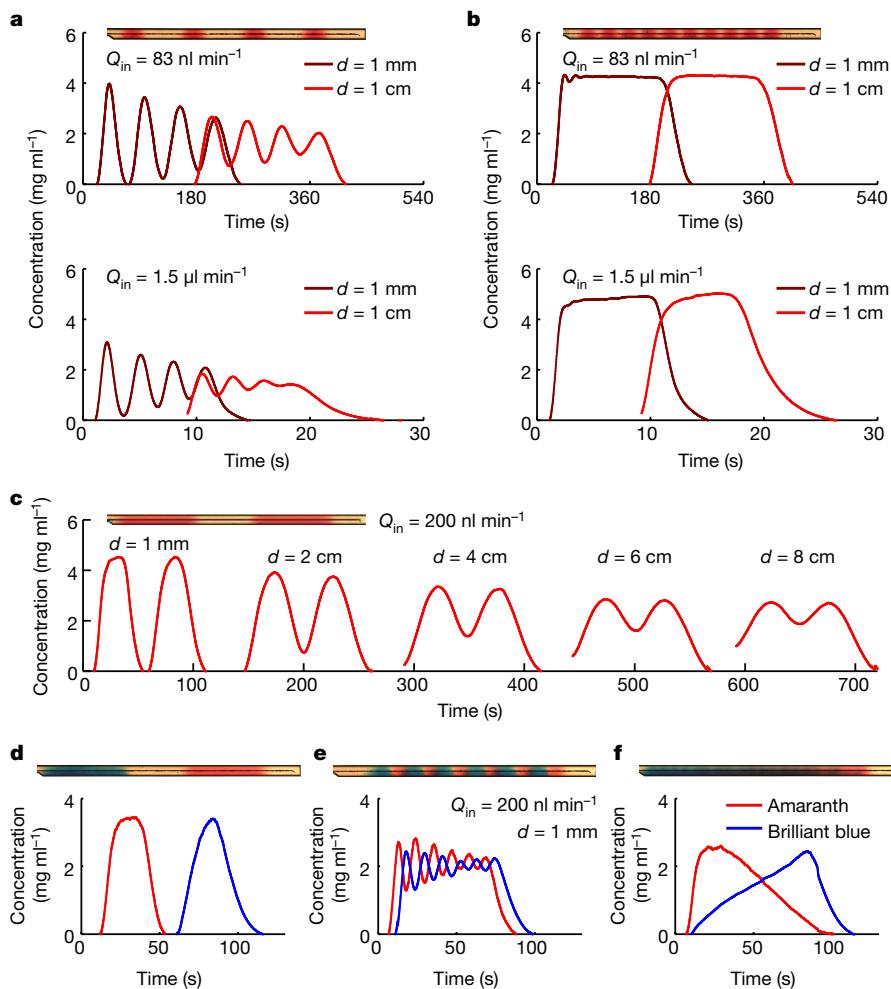
**Fig. 4 | Reagent pulse shaping and generation of complex concentration profiles.** These mean concentration profiles of reconstituted reagent were obtained using various deposition patterns, flow rates, $Q_{in}$ values and measurement distances, $d$ ($n = 6$ for each plot). The experimental profiles (optical micrographs) above the plots were obtained before the liquids exited the SCM (13-mm long). **a**, Well separated reagent spots disperse and acquire a broader concentration profile as they travel in the SCM and the downstream channel, at both the optimal flow rate, 83 nl min$^{-1}$ (top) and a fast flow rate, 1.5 μl min$^{-1}$ (bottom). **b**, Sufficiently dense spotting of reagents leads to homogenous concentration profiles. **c**, Evolution of the concentration domains of a reagent spotted at two main locations. **d–f**, Sequential delivery of two types of reagent, which were spotted such that they were well separated (**d**), co-homogenizing (**e**), or forming gradients (**f**).

G6PDH—a concentration 130 times lower than the 10% activity cut-off used in clinical tests to diagnose inherited G6PDH deficiency[24], which affects 400 million people worldwide[25].

Self-coalescence is a very general concept and applicable to other biochemical assays. We use it here also to perform isothermal recombinase polymerase amplification (RPA)[26] for the detection of DNA sequences of human papilloma virus (HPV) types 16 and 18 (Fig. 5d). For this implementation, the reaction master mix (containing enzymes, nucleotides, and so on) for RPA was deposited in SCM 1, and other reagents (primers, Mg$^{2+}$ and SYBR Green dye) were deposited in SCM 2. The flow was paused for 3 min once SCM 1 was filled in order to allow complete reconstitution of the viscous RPA master mix. A self-timing SCM can also be used to delay the release of liquids (Extended Data Fig. 6 and Supplementary Information, section 2.3).

RPA is a more complicated chemical system than G6PDH quantification and requires finer optimization. For example, increasing the concentration of SYBR Green yields a better signal-to-noise ratio, but such intercalating dyes interfere with polymerase reactions and delay the onset of DNA amplification (Fig. 5e and Extended Data Fig. 7a). Similarly, amplification reactions accelerate with increasing Mg$^{2+}$ concentration (Fig. 5f and Extended Data Fig. 7b), but too much Mg$^{2+}$ can lead to unspecific amplification and high background noise owing to the formation of primer dimers. Reaction conditions can be readily mapped by spotting reagents in an SCM, as done here using step-function gradients of SYBR Green and Mg$^{2+}$ (Fig. 5d). A DNA concentration gradient in an SCM can also be used to calibrate the real-time RPA reaction for quantifying DNA (Fig. 5g and Extended Data Fig. 7c). Moreover, different sets of primers can be deposited at different locations of SCMs for multiplexed detection of multiple target sequences (Fig. 5h, i) with high specificity and reproducibility (Fig. 5j).

The minimal dispersion of reagents in SCMs enables the localization of neighbouring reactions without compartmentalization, so that complex biochemical reactions can be implemented in nanolitre volumes of liquids in a single or several combined SCMs. This capability should have far-reaching consequences for biological assays and diagnostics, but could also prove to be a game-changing technology for chemistry at the microscale, and have impacts on the synthesis and discovery of new materials and research in the life sciences.

## Online content
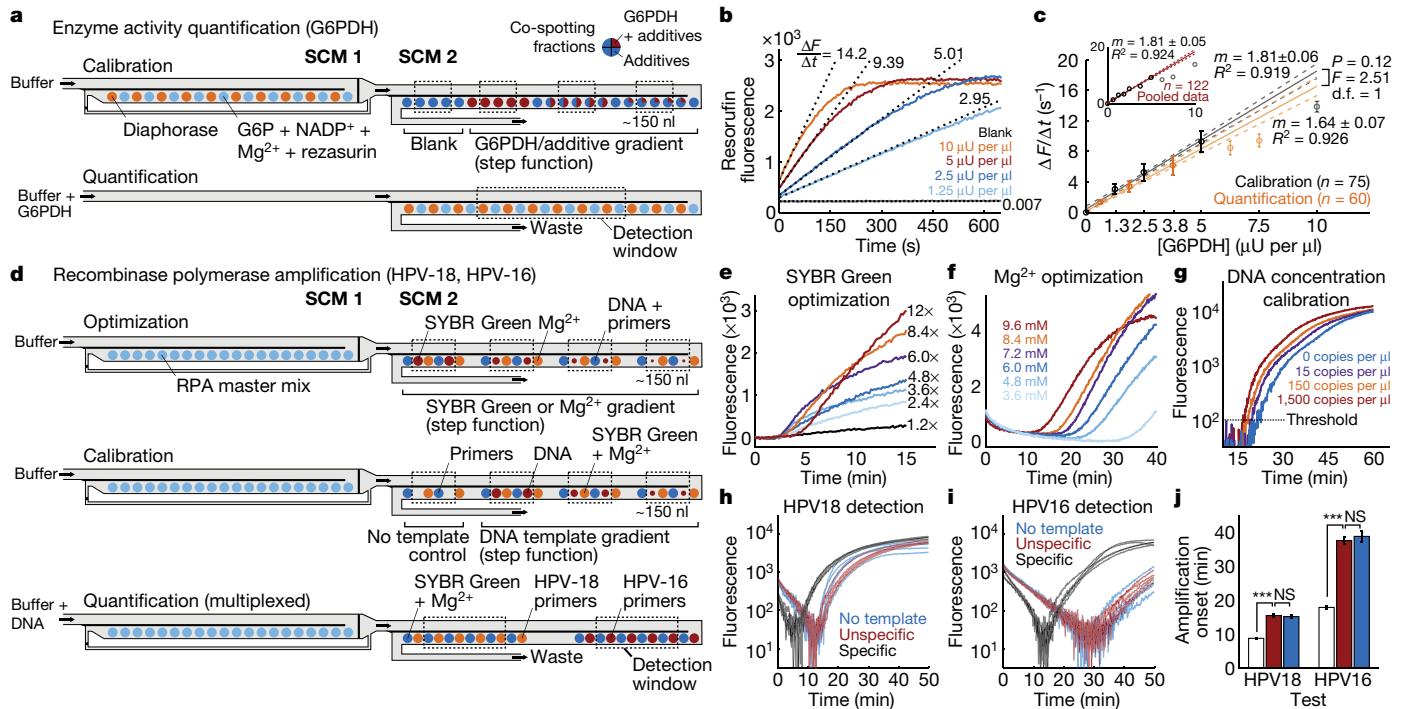
**Fig. 5 | Biochemical reactions in SCMs. a**, Implementation of a fluorometric reaction for quantifying active G6PDH. To calibrate the reaction (top), all reagents except the G6PDH analyte are deposited alternatingly into SCM 1, where they reconstitute and homogenize in a buffer before filling SCM 2. SCM 2 contains a G6PDH gradient and a counter gradient of additives. To quantify active G6PDH in a buffer (bottom), only one SCM is used. Fluorescence signals are measured over separate detection windows (dashed rectangles). Extra buffer from SCM 1 is purged at the entrance of SCM 2 using a vent/waste channel. **b**, Calibration data from a single experiment, based on the highest rate of change in resorufin fluorescence ($\Delta F / \Delta t$; dashed lines) at different G6PDH concentrations (coloured lines). **c**, Data from calibration (black) and quantification assays (orange) show a linear relationship between $\Delta F / \Delta t$ and G6PDH for low concentrations, and a saturation profile for higher concentrations (faint data points, omitted from the regression analysis), as expected for such an enzyme-concentration activity curve. The calibration and assay data are not significantly different ($n$, number of data points; $m$, slope of fit; d.f., degrees of freedom; error bars, standard deviation; calibration, ten experiments or more; quantification,

five experiments or more). The inset shows the regression analysis of calibration and assay data pooled together. **d**, Implementation of RPA using SCMs, where SCM 1 contains the RPA reaction master mix and SCM 2 contains the rest of the reagents. RPA reaction kinetics are characterized locally by measuring fluorescence over concentration gradients of SYBR Green, $Mg^{2+}$ or DNA template. A multiplexed quantification of DNA concentration can be realized by depositing primers for different templates in separate areas of SCM 2. **e, f**, Optimization data from single experiments at different SYBR Green (**e**) and $Mg^{2+}$ (**f**) concentrations, for amplifying ten copies per microlitre of HPV-18 DNA. **g**, DNA concentration calibration data from one experiment for HPV-18 DNA quantification (threshold, 100). **h, i**, Individual traces (based on five experiments) of fluorescence signals from the HPV-18 detection window (**h**) and the HPV-16 detection window (**i**) when the test was run with no template (blue), with 1,000 copies per microlitre of unmatched template (unspecific, red), or with 1,000 copies per microlitre of matching template (specific, black). **j**, Amplification onset times from different tests (white, specific; red, unspecific; blue, no template). ***$P < 0.001$; NS, not significant; error bars, standard error of the mean.

1. Yager, P. et al. Microfluidic diagnostic technologies for global public health. *Nature* **442**, 412–418 (2006).
2. Hughes, T. R. et al. Expression profiling using microarrays fabricated by an ink-jet oligonucleotide synthesizer. *Nat. Biotechnol.* **19**, 342–347 (2001).
3. Derby, B. Printing and prototyping of tissues and scaffolds. *Science* **338**, 921–926 (2012).
4. Sirringhaus, H. et al. High-resolution inkjet printing of all-polymer transistor circuits. *Science* **290**, 2123–2126 (2000).
5. Gökçe, O., Mercandetti, C. & Delamarche, E. High-content optical codes for protecting rapid diagnostic tests from counterfeiting. *Anal. Chem.* **90**, 7383–7390 (2018).
6. Hitzbleck, M. & Delamarche, E. Reagents in microfluidics: an 'in' and 'out' challenge. *Chem. Soc. Rev.* **42**, 8494–8516 (2013).
7. Fridley, G. E., Le, H. Q., Fu, E. & Yager, P. Controlled release of dry reagents in porous media for tunable temporal and spatial distribution upon rehydration. *Lab Chip* **12**, 4321–4327 (2012).
8. Beck, M. et al. On-chip sample preparation by controlled release of antibodies for simple CD4 counting. *Lab Chip* **12**, 167–173 (2012).
9. Hitzbleck, M., Gervais, L. & Delamarche, E. Controlled release of reagents in capillary-driven microfluidics using reagent integrators. *Lab Chip* **11**, 2680–2685 (2011).
10. Garcia, E., Kirkham, J. R., Hatch, A. V., Hawkins, K. R. & Yager, P. Controlled microfluidic reconstitution of functional protein from an anhydrous storage depot. *Lab Chip* **4**, 78–82 (2004).
11. Hele-Shaw, H. S. The flow of water. *Nature* **58**, 34–36 (1898).
12. Oliver, J. F., Huh, C. & Mason, S. G. Resistance to spreading of liquids by sharp edges. *J. Colloid Interface Sci.* **59**, 568–581 (1977).
13. de Gennes, P. G., Brochard-Wyart, F. & Quéré, D. *Capillary Wetting Phenomena: Drops, Bubbles, Pearls, Waves* 225–226 (Springer, 2004).
14. Howison, S. Complex variable methods in Hele-Shaw moving boundary problems. *Eur. J. Appl. Math.* **3**, 209–224 (1992).
15. Goyette, P.-A. et al. Microfluidic multipoles theory and applications. *Nat. Commun.* **10**, 1781 (2019).
16. Driscoll, T. A. & Trefethen, L. N. *Schwarz–Christoffel Mapping* (Cambridge Univ. Press, 2002).
17. Bensimon, D., Kadanoff, L. P., Liang, S., Shraiman, B. I. & Tang, C. Viscous flow in two dimensions. *Rev. Mod. Phys.* **58**, 977–999 (1986).
18. Zimmermann, M., Schmid, H., Hunziker, P. & Delamarche, E. Capillary pumps for autonomous capillary systems. *Lab Chip* **7**, 119–125 (2007).
19. Vulto, P. et al. Phaseguides: a paradigm shift in microfluidic priming and emptying. *Lab Chip* **11**, 1596–1602 (2011).
20. Safavieh, R. & Juncker, D. Capillarics: pre-programmed, self-powered microfluidic circuits built from capillary elements. *Lab Chip* **13**, 4180–4189 (2013).
21. Zhao, B., Moore, J. S. & Beebe, D. J. Surface-directed liquid flow inside microchannels. *Science* **291**, 1023–1026 (2001).
22. Weiner, A. M. Femtosecond pulse shaping using spatial light modulators. *Rev. Sci. Instrum.* **71**, 1929–1960 (2000).
23. Guilbault, G. G. & Kramer, D. N. Fluorometric procedure for measuring the activity of dehydrogenases. *Anal. Chem.* **37**, 1219–1221 (1965).
24. LaRue, N. et al. Comparison of quantitative and qualitative tests for glucose-6-phosphate dehydrogenase deficiency. *Am. J. Trop. Med. Hyg.* **91**, 854–861 (2014).
25. Cappellini, M. & Fiorelli, G. Glucose-6-phosphate dehydrogenase deficiency. *Lancet* **371**, 64–74 (2008).
26. Piepenburg, O., Williams, C. H., Stemple, D. L. & Armes, N. A. DNA detection using recombination proteins. *PLoS Biol.* **4**, e204 (2006).

## METHODS

**Chip fabrication and reagent deposition.** Silicon wafers with a 600-nm thermally grown oxide layer were processed using standard photolithography and deep reactive-ion etching (DRIE) (Extended Data Fig. 2a). Briefly, microchannels were patterned on the oxide layer using a 1.2-μm-thick AZ 6612 positive-tone photoresist and a glass/chromium photomask. Following a mild plasma cleaning to remove the photoresist residues to activate the surface, the oxide layer was etched in buffered hydrofluoric acid (BHF). The photoresist was striped in a plasma asher and a new layer of AZ 6612 photoresist was spin-coated and patterned to define CPL geometries. The silicon substrate was anisotropically etched for 30 μm using DRIE (Alcatel AMS 200). Following the removal of the photoresist layer, both microchannels and CPLs were etched for 50 μm using DRIE and the oxide layer as a hard mask. Subsequently, the oxide layer was removed in BHF and the wafer was diced. The quality of the fabrication was verified using scanning electron microscopy (Zeiss Leo 1550) and optical interferometry (Veeco Wyco NT1100).

We also fabricated SCMs in a polymeric photoresist (SU-8) as part of capillary-driven microfluidic chips, and designed SCMs with relaxed dimensions and rounded corners for fabrication using injection moulding. SCMs should also be compatible with other replication techniques involving polydimethylsiloxane (PDMS) moulding and hot embossing.

We prefer using narrow trenches (5-μm wide) to create leading barriers and a step down to create diversion barriers (Extended Data Fig. 2b–f): this implementation has the advantage of not increasing the flow resistance over the CPL, is robust, and is easily fabricated in silicon channels of any thickness using standard lithography and DRIE (Extended Data Fig. 2a). Notwithstanding, protruding leading barriers (that is, a rail-like geometry) provide better stability than trenches for the same SCM height, because the smaller liquid–vapour interface leads to a proportionally larger Laplace pinning pressure. The geometry of the diversion barrier is not critical as this barrier is challenged only when the SCM is completely filled. However, if the gap between the leading barrier and the diversion barrier is too narrow (for example, less than 200 μm in 50-μm-deep SCMs), it pins the liquid, which in turn increases the pressure on the barriers and can result in their failure.

The surface of the microfluidic chips was cleaned using an air plasma (for 2 min with a coil power of 200 W; Tepla 100-E). Within 10 min after cleaning, the chips were silanized by immersing them in a solution of 0.1% trichloro(octyl)silane (Sigma-Aldrich) in heptane (Sigma-Aldrich) for 2 min. After rinsing the chips with ethanol (Fluka) and drying them under a stream of nitrogen, solutions of erioglaucine disodium (Sigma-Aldrich, hereon referred to as brilliant blue) and amaranth (Sigma-Aldrich), each at 30 mg ml$^{-1}$ in water, were deposited in microfluidic structures using an inkjet spotter (Nano-Plotter 2.1, Gesim GmbH) equipped with a PicoTip-A piezoelectric pipette (Gesim GmbH). First, the reagents were spotted in two alignment marks (one visible in Extended Data Fig. 3) located at opposite sides of the chips. The difference between the programmed spotting coordinates and the effective location of spots was used to eliminate any misalignment. After spotting specific patterns of reagents on microfluidic chips, the chips were sealed with 3-mm-thick slabs of PDMS (Dow Corning Slygard 184). The volume of jetted droplets was measured by depositing 1,000 droplets of amaranth or brilliant blue solution at a spot on a clean surface, letting the spot dry, reconstituting the spot in 1 μl of water and measuring the concentration of the reconstitution with a spectrophotometer (Tecan Infinite M200). The volume of the droplets varied slightly from 40 pl to 60 pl between different spotting sessions, but was stable within the same session. The amount of deposited reagents was calibrated according to the droplet volume.

The patterns of reagents spotted in SCMs or control channels are detailed below. In experiments in which the dissolution in an SCM was compared with the dissolution in a control channel (Fig. 2d) and in volume scaling experiments (Fig. 3b), 5- to 15-mm-long lines comprising 25-ng spots (250-μm pitch) of amaranth were spotted. In experiments in which reagent pulse shaping was evaluated by using the amaranth dye (Fig. 4a–c), either four 100-ng spots (3-mm pitch; Fig. 4a), ten 100-ng spots (1-mm pitch; Fig. 4b) or two 3-mm-long lines comprising 25-ng spots (250-μm pitch) were spotted with 3-mm separation between the lines (Fig. 4c). In experiments in which the sequential delivery of multiple reagents was demonstrated (Fig. 4d), one 3-mm-long line comprising 15-ng spots (250-μm pitch) of amaranth and another one of brilliant blue were spotted with 3-mm separation between the lines. In experiments in which in situ mixing of multiple reagents was demonstrated (Fig. 4e), six 15-ng spots (1.5-mm pitch) of amaranth intercalated with six 15-ng spots (1.5-mm pitch) of brilliant blue were spotted. In experiments in which concentration gradients were generated (Fig. 4f), 30 spots (300-μm pitch) of amaranth gradually decreasing in mass from 15 ng to 0 ng were spotted over 30 spots (300-μm pitch) of brilliant blue gradually increasing in mass from 0 ng to 15 ng.

In the chips in which reagent pulse shaping was evaluated (Extended Data Fig. 8a), a narrow loop geometry[27] was used to minimize the effect of the turn on dispersion.

**Particle image velocimetry.** SCMs with a length of 15 mm, a width of 0.5 mm, a depth of 50 μm, and 5-μm-wide trenches were filled at a constant flow rate (active pumping, Kent Scientific Genie syringe pump, 0.5 μl min$^{-1}$) with a 10 μl min$^{-1}$ fluorescein sodium (Fluka) aqueous solution containing 0.02% fluorescent microspheres (size 4.8 μm; Thermo-Scientific Fluoro-Max). The filling was imaged using a fluorescence microscope (Nikon Eclipse 90i, Nikon DS-1QM/H) at a rate of 15 frames per second (fps), with a 60-ms exposure per frame. The displacement of particles crossing the leading barrier during the 60-ms exposure was measured manually in each frame from the particle streaks on the images. These measurements were used to calculate the velocity of the particles. The position of the measured particles was determined from the midpoint of the streaks. In order to display the velocity streamlines as in Fig. 1b, 15 subsequent frames were aligned (taking the position of the curved meniscus as reference) in an image stack, the maximum projection of the image stack was taken, and the projected image was inverted. For this, images of the particles that sediment during longitudinal flow were manually removed from each frame for visual clarity. Fiji[28] was used for data extraction and image processing.

**Optical measurements of reagent reconstitution.** Microfluidic chips having SCMs or control microchannels spotted with reagents were placed under a stereo microscope (Leica MX16). The optical zoom, lighting intensity, and colour and exposure settings of the RGB CMOS camera (Leica MC170 HD) were identical for all experiments. Before the start of the experiments, the roll and the pitch of the microfluidic chips were corrected with a custom tiltable XY stage. A syringe pump (Kent Scientific Genie) was connected to the microfluidic chips with 1/32-inch tubing (Extended Data Fig. 9a). Before each experiment, the rate of pumping was calibrated. During the experiments, microfluidic chips were filled with water at a constant flow rate (indicated by $Q_{in}$ for different sets of experiments). Videos (30 fps, 1,280 × 720 pixels) of the liquid carrying reconstituted reagents were recorded at fixed locations. Acquisition was stopped when the reagent solution was completely flushed away from the imaging area.

Video files were processed to extract the absorbance of solutions. This was done by defining a region of interest (ROI) on each video, centred on the downstream microchannel at a fixed distance (indicated by $d$ for different sets of experiments) away from the diversion barrier. The ROIs were 200-μm wide and covered the width of the microchannel (180–500 μm). The mean intensity value over time of individual video channels was extracted from the area defined by the ROIs. For experiments with only amaranth dye, the intensity values from the green channel were used for further analysis; for experiments with amaranth and brilliant blue dyes, the intensity values from the red and blue channels were used. The absorbance signal was calculated by taking the negative logarithm of the quotient of the intensity values over the mean value of signal coming only from water.

The absorbance signal was converted into concentration values using calibration curves taken for different sets of experiments. For this, solutions of dyes with known concentrations were filled into the microfluidic chips and the intensity values were processed as described above. The exponential fit to the calibration data was used to transform the absorbance values to concentration values (Extended Data Fig. 1c, d). For experiments in which both amaranth and brilliant blue dyes were used, the calibration curves and the absorbance data were used after linear spectral unmixing[29] (Extended Data Fig. 8b–e). Matlab was used for these analyses.

**General preparation for biochemical reactions in SCMs.** To implement biochemical reactions using SCMs (Fig. 5), we fabricated microfluidic chips containing SCMs with protruding CPLs (that is, rails instead of trenches) in silicon and silanized them as above. These chips had a channel depth of 100 μm and 40-μm-high rails (Extended Data Fig. 2g, h). The reagents were deposited inside SCMs using the inkjet spotter equipped with a NanoTip piezoelectric pipette (GeSiM GmbH). The droplet volume for each reagent solution was characterized as above. After deposition of reagents, the chips were sealed using a PDMS that had been passivated by exposure to a 0.2% bovine serum albumin (BSA, Sigma-Aldrich) solution in 50 mM Tris-HCl (pH 7.5) for 10 min.

**Spotting scheme for G6PDH reactions.** To quantify G6PDH activity (Fig. 5a–c), we adapted a fluorometric protocol[23] for implementation inside SCMs.

All solutions were prepared in a 20 mM Tris-HCl (pH 7.8) buffer (Fluka) containing 0.2% BSA additive. Solutions containing diaphorase and G6PDH also contained 3% trehalose (Fluka) and 1 mM tris(2-carboxyethyl)phosphine (TCEP; Thermo Scientific).

In the calibration configuration, SCM 1 (width 0.5 mm, length 33 mm, rail width 30 μm; Extended Data Fig. 2g) was patterned using two solutions. The first solution contained 0.1 U μl$^{-1}$ diaphorase (Sigma-Aldrich) and was deposited as discrete spots over a distance of 30 mm with a 1-mm pitch. Each spot was formed using four inkjet-deposited droplets (with a total volume of roughly 2.5 nl). The second solution contained the substrates and co-factors, namely 200 mM MgCl$_2$ (Sigma-Aldrich), 40 mM G6P (Sigma-Aldrich), 2 mM NADP$^+$ (Sigma-Aldrich) and 0.2 mM rezasurin (Sigma-Aldrich). The substrate solution was also deposited using a 1-mm pitch along 30 mm, but with a 0.5-mm shift relative to the diaphorase

spots, and with each spot formed using eight dispensed droplets (approximately 4.5 nl).

SCM 2 (width 0.5 mm, length 30 mm) was patterned so as to have three regions with different G6PDH concentrations and one region without any G6PDH for blank measurements. Each region consisted of 5-mm-long lines of spots with a 0.5-mm pitch, and was separated from another region by 2 mm. In different experiment sets, either one (585 pl), two or four droplets of solution containing $0.52 \times 10^{-4}$ U $\mu l^{-1}$ G6PDH (Sigma-Aldrich), or two, four or eight droplets of the same solution, were deposited to form regions with increasing G6PDH concentration. In order to achieve comparable reaction kinetics from different regions, we needed to compensate for additives in the G6PDH solution that were also deposited at different amounts. This was achieved by spotting a solution containing only the additives over the spots of G6PDH to adjust the total deposited amount per spot to four or eight droplets.

In the quantification configuration, we used only SCM 2, and patterned it identically to SCM 1 in the calibration configuration.

**Spotting scheme for recombinase polymerase amplification.** We implemented isothermal recombinase polymerase DNA amplification[26] (RPA) using SCMs (Fig. 5d–j and Extended Data Fig. 7a–c) by adapting and applying the reagents from a commercially available kit (TwistDx TwistAmp Basic). Quantitative synthetic HPV-18 DNA (ATCC VR-3241SD) and quantitative synthetic HPV-16 DNA (ATCC VR-3240SD) were purchased and used in the amplification reactions as the template. Primer sequences were taken from the literature[30] and the primers were purchased (Integrated DNA Technologies). The following primers were used in the experiments (5′-3′):

HPV-16 forward: TTGTTGGGGTAACCAACTATTTGTTACTGTT
HPV-16 reverse: CCTCCCCATGTCTGAGGTACTCCTTAAAG
HPV-18 forward: GCATAATCAATTATTTGTTACTGTGGTAGATACCACT
HPV-18 reverse: GCTATACTGCTTAAATTTGGTAGCATCATATTGC

In all experiments, SCM 1 (width 1 mm, length 28 mm, rail width 75 μm, rail split with a 15-μm-wide trench; Extended Data Fig. 2h) was patterned with the 2× RPA master mix, prepared by reconstituting the pellet from the kit in 1 mM TCEP solution. The RPA master mix was deposited along a 25-mm line of spots, with a 1-mm pitch, and with each spot formed using 100 droplets (approximately 50 nl).

SCM 2 (width 0.5 mm, length 30 mm) was patterned with all other reagents necessary for the RPA reaction. For optimization of the SYBR Green concentration, a DNA-Mg mix consisting of 100 copies per microlitre of HPV-18 DNA, 140 mM magnesium acetate (MgOAc; Fluka), 5 μM of each HPV-18 primer, 3% trehalose and 0.2% BSA was deposited as a 28-mm line of spots with a 0.5-mm pitch, and with each spot formed using five droplets (approximately 3 nl). Over these spots, a step-function-like gradient of SYBR Green was created by depositing a 50× SYBR Green (Invitrogen) solution containing 0.2% BSA along 4-mm segments having a 0.5-mm pitch, with the segments formed either with one (roughly 600 pl), two, three, four, five, seven or ten droplets, without additional spacing between segments.

For optimization of the Mg$^{2+}$ concentration, a 24-mm line of spots of DNA solution, containing 250 copies per microlitre of HPV-18 DNA, 12.5 μM of each HPV-18 primer and 3% trehalose, was patterned in SCM 2 with a 0.5-mm-pitch and with each spot formed by two droplets (approximately 1.2 nl). A similar second line, which was formed using spots (two droplets, approximately 1.2 nl) of 125× SYBR Green, 75 mM MgOAc, 3% trehalose and 0.2% BSA, was patterned with a 0.25-mm shift from the first line. Over the second line, a step-function-like gradient of Mg$^{2+}$ was formed by depositing 50 mM MgOAc solution on 4-mm segments having a 0.5-mm pitch, and with the segments formed either by zero, one (roughly 600 pl), two, three, four or five droplets, and no additional spacing between segments.

For calibration of DNA concentration to amplification onset time, two 18-mm lines each with a 0.5-mm pitch and spots of two droplets (roughly 1.2 nl) were patterned with a 0.25-mm offset between them. The first line consisted of spots of 125× SYBR Green, 95 mM MgOAc, 3% trehalose and 0.2% BSA. The second line consisted of spots of 95 mM MgOAc, 8.4 μM of each HPV-18 primer and 3% trehalose. Over the second line, a step-function-like gradient of HPV-18 DNA was patterned by depositing two droplets (roughly 1.2 nl) of DNA solution containing 3% trehalose along 4-mm segments with a 0.5-mm pitch, with the segments spotted from different DNA solutions having 250, 2,500 or 25,000 copies per microlitre of HPV-18 DNA. No calibration was performed for HPV-16.

For multiplexed detection of HPV-16 and HPV-18 DNA, SCM 2 was patterned similarly to the spotting scheme for DNA calibration, but without any DNA spotted. Additionally, one half of SCM 2 was patterned using HPV-16 primers and the other half using HPV-18 primers.

**Biochemical reactions in SCMs, data acquisition and analysis.** To start the biochemical reactions, spotted and sealed chips were filled only with buffer, or with buffer containing G6PDH or DNA under active pumping at varied flow rates and at room temperature (Extended Data Fig. 9a). Each chip carried elements for five

experiments. Experiments were run one after another. In all experiments, SCM 1 was filled at a rate of 1.0 μl min$^{-1}$, and SCM 2 at 0.5 μl min$^{-1}$ during longitudinal flow and 3.5 μl min$^{-1}$ during self-coalescing flow. For RPA reactions, the flow was paused for 3 min after SCM 1 was filled for the rehydration and reconstitution of the RPA master mix.

For calibration of the G6PDH reaction, the chips were filled with 50 mM Tris-HCl (pH 7.8) solution. For G6PDH quantification experiments, chips were manually filled using a micropipette (Extended Data Fig. 9b) with 50 mM Tris-HCl (pH 7.8) solutions containing 0.625–7.5 μU μl$^{-1}$ G6PDH, 0.2 mM TCEP, 0.6% trehalose and 0.04% BSA.

For calibration of the RPA reaction and for optimization of SYBR Green and MgOAc concentrations, the chips were filled with 25 mM Tris-HCl (pH 7.9) solution containing 100 mM potassium acetate (Sigma-Aldrich), 5.4% polyethylene glycol (molecular weight 35 kDa; Sigma-Aldrich) and 0.05% Tween-20 (Fluka). For the multiplexed detection of HPV-16 and HPV-18, the chips were filled with the same solution additionally containing 1,000 copies per microlitre of HPV-16 or HPV-18 DNA.

After the chips were filled, they were swiftly transferred to a microtitre plate reader (Tecan Infinite M200) using a custom-made aluminium adaptor (Extended Data Fig. 9c). While the reactions ran, the reader scanned along SCM 2 and measured the fluorescence signal every millimetre. G6PDH reactions were run at 25 °C. The conversion of rezasurin to resorufin was measured at 560-nm excitation and 595-nm emission using a sampling time interval of 5 s. RPA reactions were run at 33–36 °C. SYBR Green fluorescence was measured using 490-nm excitation and 525-nm emission wavelengths every 10 s. Subsequent analyses were run only on the measurements taken at detection windows (roughly 150 nl, determined by the point spread function of the excitation beam of the plate reader; full width at half maximum roughly 3 mm as measured), centred over regions with different concentrations of G6PDH, SYBR Green, Mg$^{2+}$ or DNA (for optimization and calibration), or at detection windows centred over regions with G6PDH assay reagents or primers (for quantification).

The kinetics of the G6PDH reaction was characterized from the fastest rate of fluorescence change ($\Delta F / \Delta t$) for each reaction at different G6PDH concentrations[31]. The quality of the calibration curve was assessed using linear regression (LinearModel class in Matlab). To evaluate whether calibration and quantification experiments agreed, we applied analysis of covariance (ANCOVA; 'aoctool' function in Matlab).

The amplification onset of RPA reactions was extracted first by subtracting the background from the signal and later by applying a threshold to the logarithm of the data. The quality of the calibration curve was assessed using linear regression (LinearModel class in Matlab). To test whether detection of HPV-16 or HPV-18 DNA produced signals with significantly earlier onset times than unspecific amplification, we applied a two-sample $t$-test ('ttest2' function in Matlab).

**Finite element method for characterization of self-coalescing flow.** The finite-element method (FEM) platform COMSOL Multiphysics was used to model the SCM and simulate 3D laminar and 2D potential flows. Both geometries were defined using COMSOL's CAD tools. The parameters used in the models are shown in Extended Data Table 1. The 3D laminar and 2D potential flows were respectively defined using COMSOL's 'laminar flow' and 'mathematics' modules—that is, incompressible Navier–Stokes equation and Laplace equation. In both cases, the problem is solved in the reference frame of the moving meniscus (the $Z*$ domain in Supplementary Information section 1) and plotted in the laboratory reference frame (the $Z$ domain) with the transformation $z* = z + Ut$. This equivalence specifically implies the invariance property of the Navier–Stokes equation under a Galilean transformation. These equations were both solved using the built-in steady-state fully coupled solver. Elements were manually refined on the boundary of the meniscus to obtain sufficient numerical accuracy. Parametric sweeps were used to study the impact of the total width, contact angle and flow rate on the velocity field. The orthogonal velocity component was taken from a line going through the SCM along the $x$ axis at different $y$ positions and at mid-height in the Navier–Stokes case.

## Data availability

The data that support the findings of this study are available from the corresponding authors on request.

## Code availability

The code for the Schwarz–Christoffel mapping of self-coalescence (using Matlab), as well as the code for reagent pulse shaping (using Maple), is available on request from T.G.

27. Paegel, B. M., Hutt, L. D., Simpson, P. C. & Mathies, R. A. Turn geometry for minimizing band broadening in microfabricated capillary electrophoresis channels. *Anal. Chem.* **72**, 3030–3037 (2000).

28. Schindelin, J. et al. Fiji: an open-source platform for biological-image analysis. *Nat. Methods* **9**, 676–682 (2012).
29. Zimmermann, T. Spectral imaging and linear unmixing in light microscopy. *Adv. Biochem. Eng. Biotechnol.* **95**, 245–265 (2005).
30. Seaman, W. T. et al. Detection and quantitation of HPV in genital and oral tissues and fluids by real time PCR. *Virol. J.* **7**, 194 (2010).
31. Glock, G. E. & McLean, P. Further studies on the properties and assay of glucose 6-phosphate dehydrogenase and 6-phosphogluconate dehydrogenase of rat liver. *Biochem. J.* **55**, 400–408 (1953).

**Author contributions** E.D. conceived the research. O.G. discovered self-coalescing flows, designed SCMs and experiments, and conducted the experiments. T.G. developed the theoretical framework and contributed to the experimental design. S.C. performed the conformal mapping calculations and the numerical simulations, and contributed to the theoretical framework. O.G., S.C. and T.G. analysed the data. Y.T. fabricated microfluidic chips and documented them using scanning electron microscopy. O.G., T.G. and E.D. wrote the manuscript. All authors discussed and provided input to the manuscript.

**Competing interests** The authors declare no competing interests.

**Additional information**
**Supplementary information** is available for this paper at https://doi.org/10.1038/s41586-019-1635-z.
**Correspondence and requests for materials** should be addressed to T.G. or E.D.
**Peer review information** *Nature* thanks Piotr Garstecki and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.
**Reprints and permissions information** is available at http://www.nature.com/reprints.
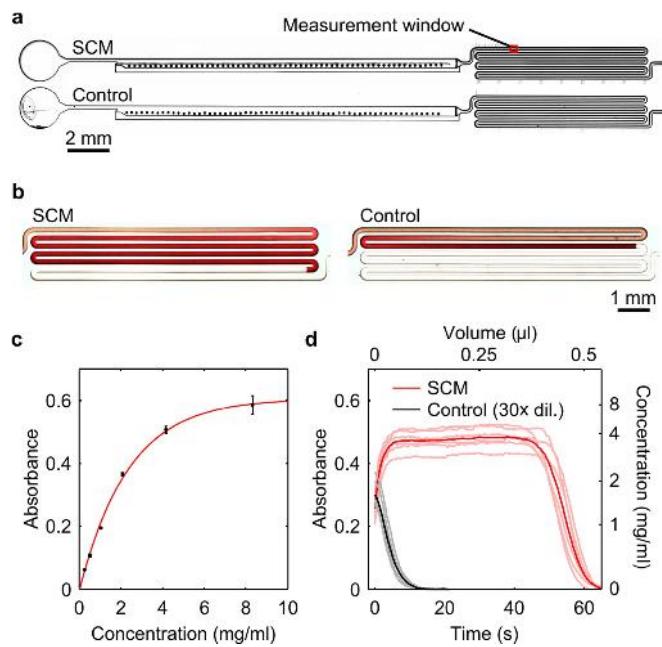
**Extended Data Fig. 1 | Characterization of reagent dissolution in SCMs and control experiments. a**, Bright-field microscope images of microfluidic chips in silicon, each chip being composed of a rounded inlet (left), an SCM (middle), and a meandering channel (right) used for conveniently measuring the concentration profile of the solution exiting the SCM. **b**, Amaranth solutions from an SCM (left) and in a control microchannel (right) are readily visualized using the meandering channels. **c**, Calibration curve for quantification of the concentration of reconstituted amaranth ($n = 4$ for each concentration). Error bars represent standard error of the mean. **d**, Concentration profiles of amaranth solution reconstituted in SCMs or control microchannels ($Q_{in} = 500$ nl min$^{-1}$). Means of individual acquisitions are displayed with dark coloured lines and individual acquisitions with a lighter shading. Amaranth solution was inkjet spotted into SCMs at 100 ng mm$^{-1}$ (250-μm pitch). The reagent accumulation in the control was so strong that the amaranth solution was diluted ×30 before deposition to keep the absorbance signal in the dynamic range of the camera. The concentration profile from control experiments shown in Fig. 2d is scaled up from this diluted signal.
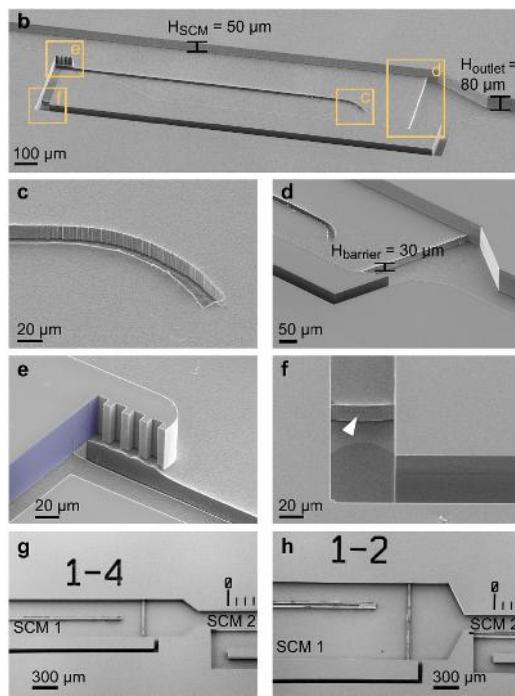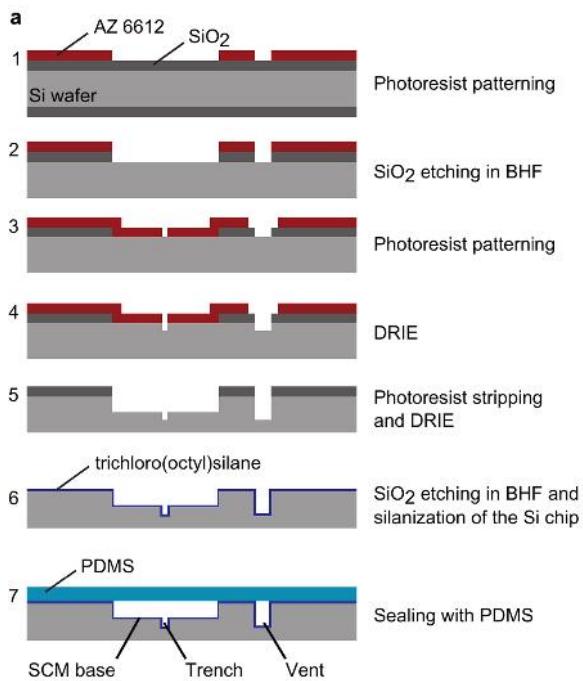
**Extended Data Fig. 2 | Fabrication steps for SCMs with depressed barriers, and scanning electron microscopy images of representative SCMs. a**, Silicon wafers are processed by standard photolithography and multiple steps of deep reactive-ion etching (DRIE). Specifically, a 1.2-$\mu$m-thick AZ 6612 photoresist layer is patterned (step 1) to mask the $SiO_2$ layer during etching in buffered hydrofluoric acid (BHF; step 2). A new layer of AZ 6621 is patterned (step 3), exposing the areas that are etched to form the trenches during the first DRIE step (step 4). Later, using the patterned $SiO_2$ layer as the mask, a second DRIE step forms the microchannels while preserving the trenches (step 5). After dicing, the chips are cleaned and silanized in trichloro(octyl)silane solution (step 6). After reagent deposition, chips are sealed with a layer of polydimethylsiloxane (PDMS;

step 7). **b**, Image of an SCM fabricated in silicon, with yellow frames highlighting particular regions shown below. **c**, The leading barrier with a trench geometry. Here, for visual clarity, the trench width is 20 $\mu$m, but 5-$\mu$m-wide trenches were used in experiments. The curved end of the leading barrier facilitates the initiation of self-coalescing flow. **d**, Diversion barrier with a step-down geometry. **e**, Raised vertical CPLs at the entrance of the SCM help to keep the meniscus away from the side wall coloured in blue. **f**, The barrier at the vent entrance (arrowhead) ensures that liquid does not enter the vent. The depths of the SCM ($H_{SCM}$), outlet ($H_{outlet}$) and barriers ($H_{barrier}$) are shown on the images. **g**, **h**, Scanning electron microscopy (SEM) images of the devices used in G6PDH reactions (**g**) and in RPA reactions (**h**). (The numbers 1-2 and 1-4 are device numbers.).
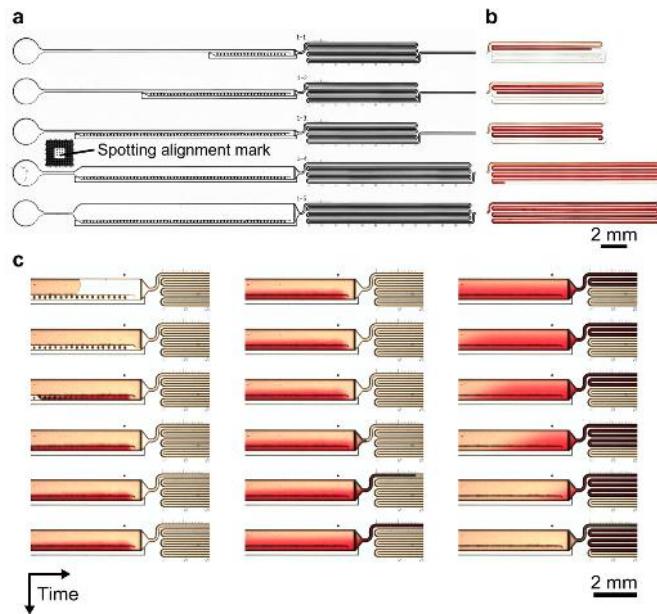
**Extended Data Fig. 3 | SCM volume scaling experiments. a**, Bright-field microscope images of devices fabricated in silicon (numbered 1-1 to 1-5) in which volume scaling experiments were performed. As discussed in Fig. 1d, the velocity of the self-coalescing flow decays more quickly in narrower than in wider SCMs. In order to maintain a quick decay of the self-coalescing flow velocity in wide SCMs, the leading barrier is simply shifted towards the area in which reagents are deposited. Amaranth is deposited at 100 ng mm$^{-1}$ (250-μm pitch). A spotting alignment mark is used to align the inkjet spotter head with the microfluidic chips for precise targeting of the deposition location. **b**, Bright-field microscope images showing the amaranth solution in the meandering channels, reconstituted in the SCMs shown in panel **a**. **c**, Time series of bright-field microscope images showing the reconstitution of amaranth in a 1-mm-wide SCM ($Q_{in} = 500$ nl min$^{-1}$). Typically, in wide SCMs (wider than 0.5 mm), lateral homogenization of the reconstituted reagent is complete at the narrow outlet of the SCM (as seen here), and can be enhanced, if necessary, by adding a Dean vortex mixer, a chaotic mixer, or a simple meandering channel. Larger volumes of solution with reconstituted reagents can also be achieved using an array of parallel SCMs. SCMs are easily scaled; however, making them longer decreases the maximum filling flow rate because the pressure at the inlet needs to be lower than the Laplace pressure over the CPL.
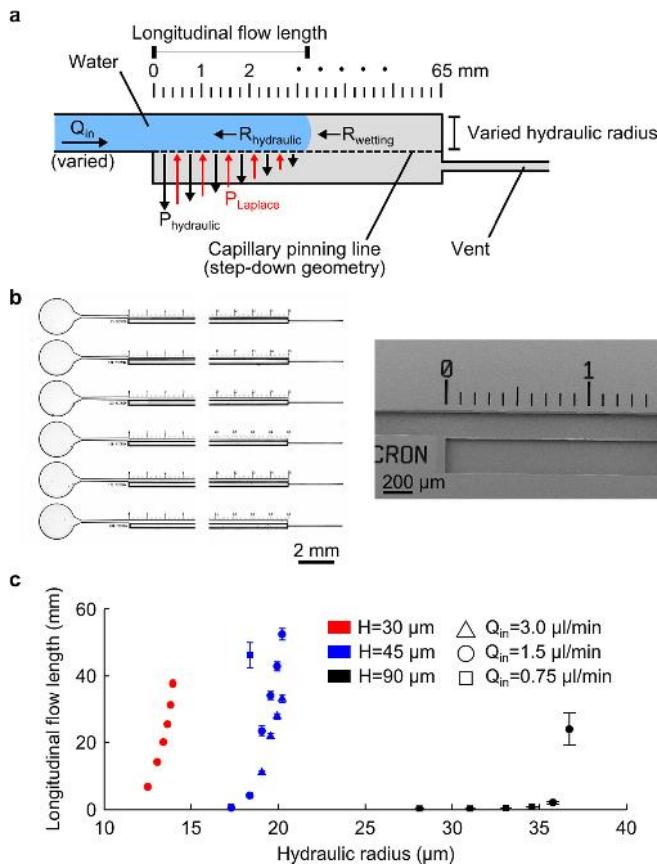
**Extended Data Fig. 4 | Assessment of the stability of the leading barrier.**
**a**, During longitudinal flow, the hydraulic pressure ($P_{hydraulic}$) increases as the hydraulic resistance ($R_{hydraulic}$) increases with the distance travelled by the liquid. The leading barrier is able to preserve the longitudinal flow as long as the pinning forces ($P_{Laplace}$) can balance the increasing hydraulic pressure and the pressure exerted by the resistance to wetting ($R_{wetting}$). In order to investigate this effect, the width ($W = 75, 100, 125, 150, 175$ or $200\,\mu m$) and the height ($H = 30, 45$ or $90\,\mu m$) of the longitudinal flow area and the filling flow rate ($Q_{in} = 0.75, 1.5$ or $3.0\,\mu l\,min^{-1}$) were varied. During experiments, the maximum distance along which the liquid remains pinned was measured. **b**, Bright-field (left) and SEM (right) images of devices in which stability-assessment experiments are performed. **c**, Experimental data support theoretical predictions that: first, a smaller liquid–vapour interface can bear a higher hydraulic pressure; second, the hydraulic radius of the longitudinal flow area needs to be large in order to fill deeper structures or for high filling flow rates; and third, the flow rate challenges the stability of the leading barrier. Data points are means of experiments ($n = 8$ or more; error bars represent standard error of the mean). Conditions in which the liquid was not pinned at the CPL (longitudinal flow length $= 0$ mm) or the liquid travelled to the end of the test device (longitudinal flow length $= 65$ mm) are excluded from the plot.
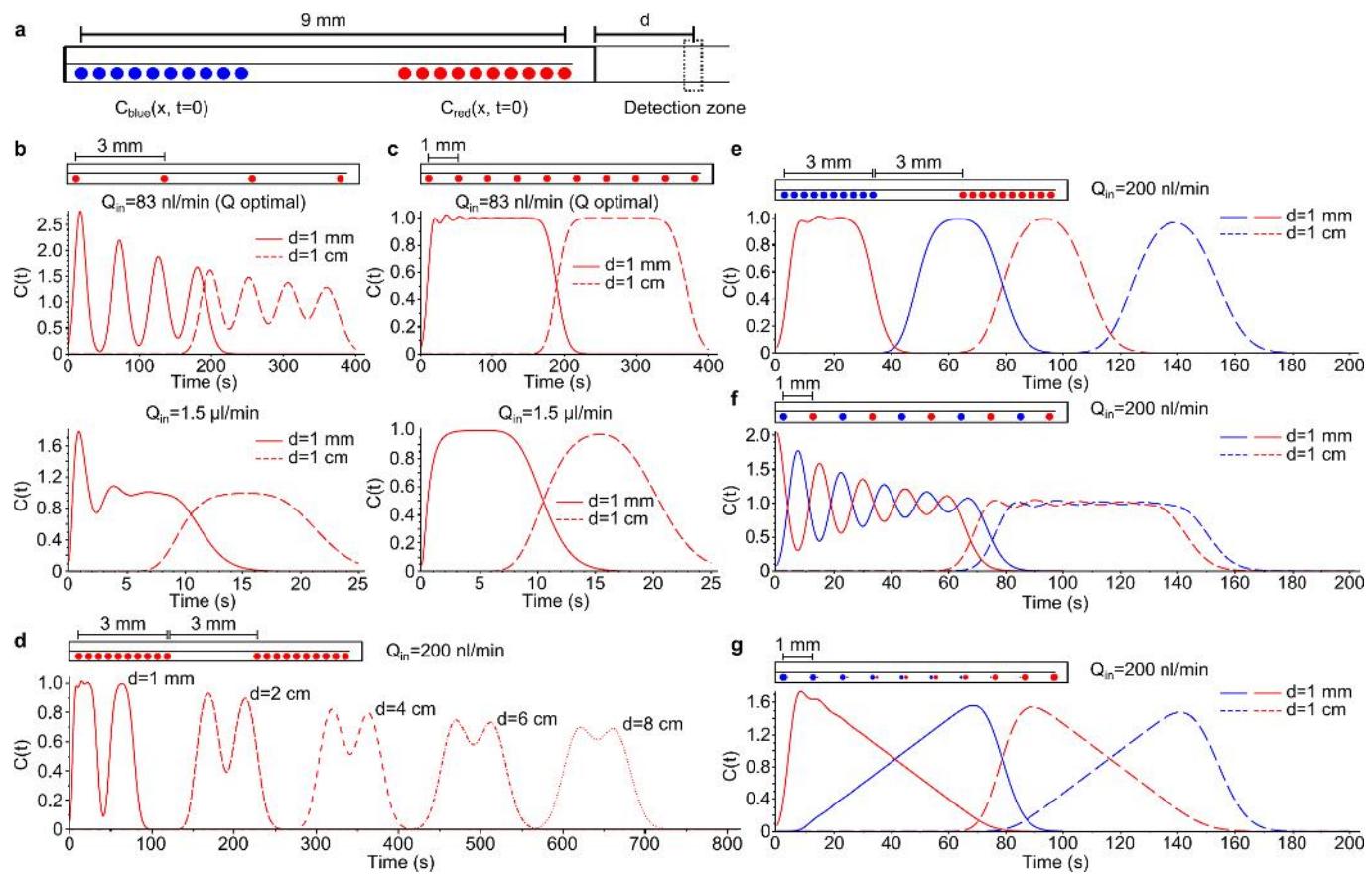
**Extended Data Fig. 5 | Predicted complex concentration profiles from diffusion–dispersion models. a**, Illustration of the experimental method for analysing dispersion in SCMs and in downstream channels at distance $d$. C, concentration. **b–g**, Predicted concentration profiles for experiments shown in Fig. 4 ($C(t)$ in arbitrary units). **b, c**, Results of controlling the spotting pitch to generate spiked (**b**) or flat (**c**) concentration pulses at optimal (Pé = 5.6; top panels) and large (Pé = 100; lower panels) Péclet numbers, at constant spotted reagent mass. **d**, Signal decorrelation over time due to Taylor–Aris dispersion for several positions of the detection zone (Pé = 13). **e**, Pulsing of two spatially resolved different reagents (Pé = 13). **f**, Pulsing of a flat profile of two mixed reagents by alternated spotting and flow control (Pé = 13). **g**, Pulsing of two linear concentration gradients to a detection zone (Pé = 13).
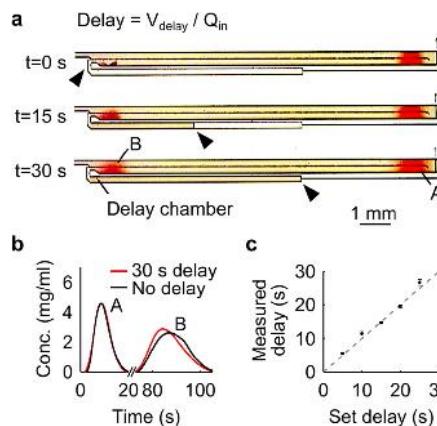
**Extended Data Fig. 6 | Release of reconstituted reagents from an SCM with time delays.** A chamber of volume $V_{delay}$ at the entrance of the vent adds a delay before a liquid leaves the SCM. **a**, Time series of images showing filling of the delay chamber while reconstituted amaranth diffuses. Reagents in area A, but not in area B, have additional time to reconstitute/diffuse while the rest of the SCM fills ($t = 0$ s). If needed, using a delay chamber gives additional time for the reconstitution/diffusion of the B reagents. Arrowheads mark the liquid filling front in the delay chamber. **b**, Effect of delay on the concentration profile of points A and B. Without any delay, point B disperses more than it does with a delay, and its profile is broader ($n = 9$; $Q_{in} = 200$ nl min$^{-1}$). **c**, Measured delay provided by different delay chambers. The dashed line is the identity line; error bars represent standard error of the mean; $n = 6$; $Q_{in} = 200$ nl min$^{-1}$.
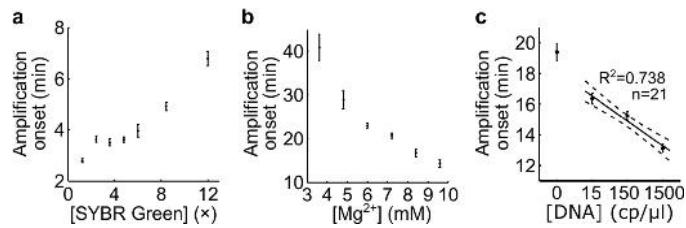
**Extended Data Fig. 7 | Optimization and calibration of RPA reactions in SCMs. a**, **b**, Optimization data for SYBR Green concentration (**a**; four experiments) and $Mg^{2+}$ concentration (**b**; five experiments) for amplification of ten copies per microlitre of HPV-18 DNA. **c**, DNA concentration calibration (seven experiments) for HPV-18 DNA quantification. Plots show amplification onset time extracted from SYBR Green fluorescence signals. Error bars represent standard errors of the mean.
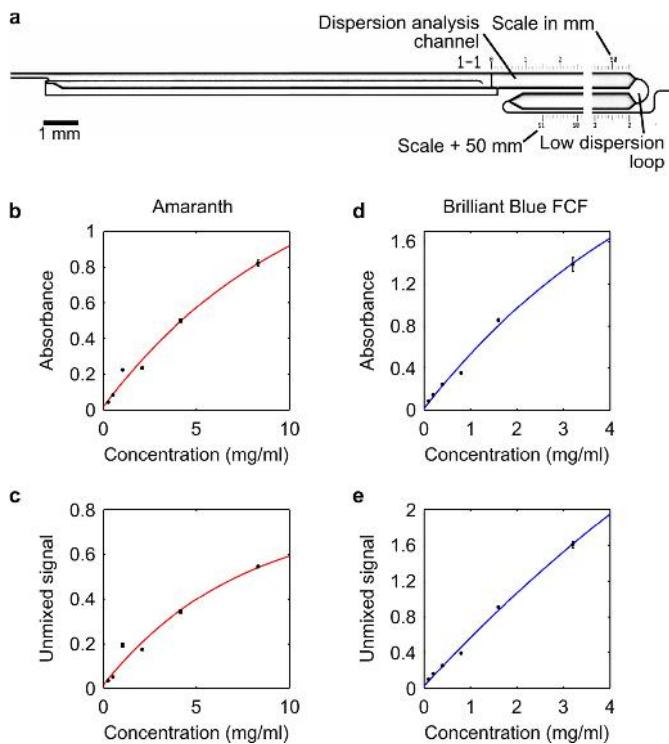
**Extended Data Fig. 8 | Characterization of dissolution of multiple reagents in SCMs. a**, Bright-field image of the device in which reagent pulse shaping and complex profile generation experiments were performed. The diversion barrier of these SCMs has a trench geometry in order to keep the dimensions of the dispersion analysis channel identical to the dimensions of the SCM. **b**, Absorbance calibration curve, and, **c**, the calibration curve after linear spectral unmixing of the amaranth signal. **d**, **e**, As for panels **b**, **c**, but for the brilliant blue signal. $n = 6$ for each data point. Error bars represent standard errors of the mean.
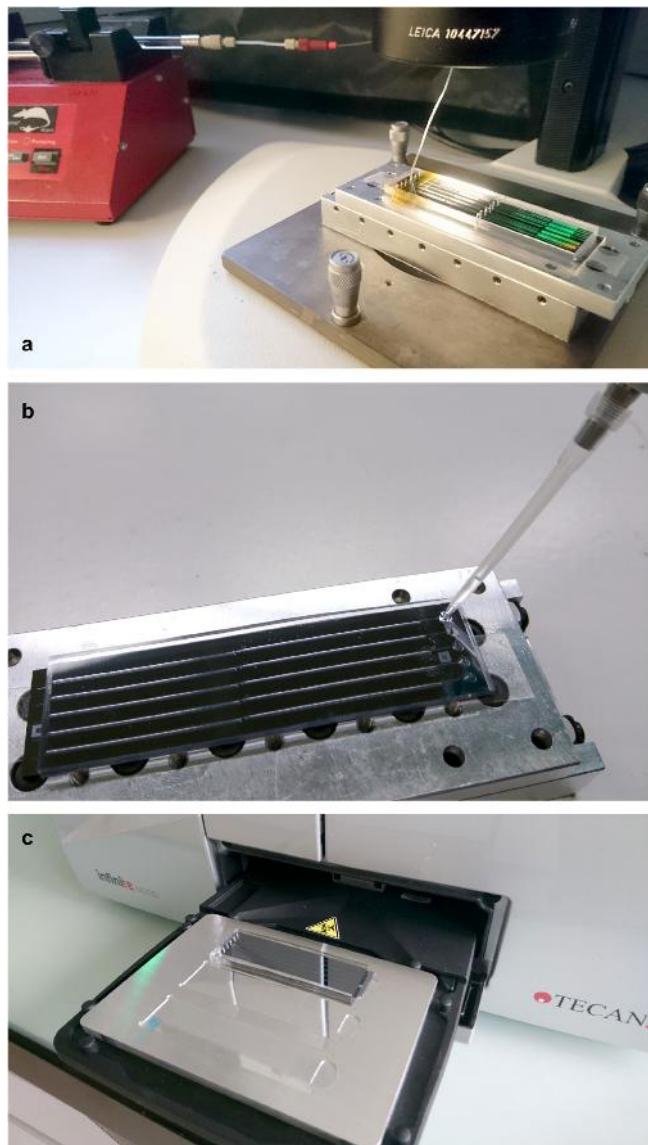
**Extended Data Fig. 9 | Experimental setup. a**, The setup for filling SCMs and documenting the dissolution of food dyes. **b**, Filling SCMs manually using a micropipette for G6PDH quantification experiments. **c**, Measuring the fluorescence signal from G6PDH and RPA reactions using a microtitre plate reader fitted with a custom-made aluminium adaptor.

**Extended Data Table 1 | The parameters used in simulations**

| Variable | Description | Value |
|:---:|:---|:---|
| $H$ | Channel height | 50 μm |
| $W$ | Channel total width | 500 μm |
| $L$ | Channel length | 1 cm |
| $\theta_w$ | Contact angle (experimentally measured) | varies (default: 116°) |
| $Q$ | Flow rate | 0.5-1.5 μl/min |
| $\rho_{H_2O}$ | Water density at ~20°C | 1 kg/m$^3$ |
| $\eta_{H_2O}$ | Water dynamic viscosity at ~20°C | 1 mPa·s |

# LETTER

# Constraints on global mean sea level during Pliocene warmth

Oana A. Dumitru[1,6], Jacqueline Austermann[2,6], Victor J. Polyak[3], Joan J. Fornós[4], Yemane Asmerom[3], Joaquín Ginés[4], Angel Ginés[4] & Bogdan P. Onac[1,5]*

**Reconstructing the evolution of sea level during past warmer epochs such as the Pliocene provides insight into the response of sea level and ice sheets to prolonged warming[1]. Although estimates of the global mean sea level (GMSL) during this time do exist, they vary by several tens of metres[2–4], hindering the assessment of past and future ice-sheet stability. Here we show that during the mid-Piacenzian Warm Period, which was on average two to three degrees Celsius warmer than the pre-industrial period[5], the GMSL was about 16.2 metres higher than today owing to global ice-volume changes, and around 17.4 metres when thermal expansion of the oceans is included. During the even warmer Pliocene Climatic Optimum (about four degrees Celsius warmer than pre-industrial levels)[6], our results show that the GMSL was 23.5 metres above the present level, with an additional 1.6 metres from thermal expansion. We provide six GMSL data points, ranging from 4.39 to 3.27 million years ago, that are based on phreatic overgrowths on speleothems from the western Mediterranean (Mallorca, Spain). This record is unique owing to its clear relationship to sea level, its reliable U–Pb ages and its long timespan, which allows us to quantify uncertainties on potential uplift. Our data indicate that ice sheets are very sensitive to warming and provide important calibration targets for future ice-sheet models[7].**

Accurate predictions of future sea-level change hinge on our understanding of how ice sheets respond to changes in global temperature. To understand ice-sheet stability under prolonged warming (such as if the current level of temperature increase continues), we can use reconstructed sea level during past periods when Earth's climate was warmer than today[1]. The Pliocene epoch (5.33 to 2.58 million years ago, Ma) was the most recent extended global warm period immediately preceding the inception of the high-magnitude glacial/interglacial variations of the Pleistocene[8]. The mid-Piacenzian Warm Period (MPWP), an interval during the Late Pliocene (3.264 to 3.025 Ma), has been used as an analogue for future anthropogenic warming since atmospheric $CO_2$ conditions were comparable to present-day values (~400 ppm)[9] and estimated global mean temperatures were elevated by 2–3 °C relative to the pre-industrial period[5].

Oxygen isotope ratios from benthic foraminifera[10] paired with deep ocean temperature estimates have been used to approximate ice-volume-equivalent GMSL changes over the Pliocene[11,12]. While invaluable, these approaches are limited by uncertainties in the methodology and a number of factors (for example, post-burial diagenesis, long-term changes in seawater chemistry and salinity) that are poorly constrained and may bias the sea-level estimates[3]. Field mapping of palaeoshorelines has been a complementary approach and has provided several local reconstructions of Pliocene sea level[13]. Local estimates also exist for the Strait of Gibraltar, where they are based on the marginal basin residence time method and measurements of planktonic foraminifera[2]. Local estimates of sea level can vary considerably from the GMSL owing to processes such as glacial isostatic adjustment (GIA)[14] and dynamic topography[15], which can be corrected for, but have substantial

uncertainties. Lastly, GMSL estimates have also been derived from climatically driven ice-sheet simulations[4,7,16], but these vary notably as a result of model uncertainties. All these challenges have led to considerable disagreement between estimates of the GMSL during the MPWP, with values ranging from about 10 m to over 50 m above present sea level (m.a.p.s.l.)[1,2,12,17,18]. This disparity hinders the assessment of past and future ice-sheet stability[7].

Here we present Pliocene sea-level data from Coves d'Artà in the western Mediterranean (Mallorca, Spain; Fig. 1a, b) that are based on U–Pb absolute-dated phreatic overgrowths on speleothems (POS). POS offer several important benefits over other Pliocene sea-level indicators since they store all information needed for a meaningful sea-level index point: (1) precise spatial geographic positioning, (2) accurate elevation, (3) clear indicative meaning (their growth covers the full tidal range, thus having an explicit relation to past sea level; see Methods), and (4) an absolute age (since the crystalline aragonite/calcite often contains suitable uranium concentrations for robust dating[19]). POS are primarily precipitated in caves, at the water/air interface as $CO_2$ degasses from brackish cave pools. The water table in these caves is, and was in the past, coincident with sea level, given that the caves are at most 300 m away from the coast and the karst topography is low. Six POS levels have been identified at elevations from 22.6 to 31.8 m.a.p.s.l. (uncertainties in the elevation measurement and the indicative range are less than 1 m; Fig. 1c, Table 1). We interpret these levels as still stands, that is, corresponding to periods of time during which sea level has been stable long enough to allow the precipitation of carbonate overgrowths. This could occur during a sea-level highstand, lowstand or intermediate stand. The palaeo sea level still stands are distinctly delineated by POS that occur either as overgrowths covering cave walls and pre-existing flowstones (top and bottom insets of Fig. 1d) or are standalone structures on stalactites and stalagmites (middle inset of Fig. 1d). Based on 70 U–Pb analyses, the geochronology of these POS yielded ages between 4.39 ± 0.39 Ma and 3.27 ± 0.12 Ma (Table 1; see Methods). These are unique because radiometric-dated records of Pliocene sea level are entirely independent of orbitally tuned chronologies.

To infer the GMSL from these local observations, the POS elevations are corrected for GIA, which is the viscoelastic adjustment of the solid Earth, its gravity field, and rotation axis to changes in the ice and ocean load. The GIA correction is the deviation of local sea level from the global mean and we calculate this correction using a gravitationally self-consistent sea-level formulation[20] paired with three GMSL scenarios[2,4,10] and a suite of viscoelastic Earth structures (see Methods). To calculate the GMSL we assume a fixed oceanic area and consider total ice-volume change, not only ice above the flotation level. Long-term deformation at passive margins due to sediment loading[21] or dynamic topography[15] can further contribute to local sea-level changes. Since model predictions of these processes have high uncertainties, we estimate bounds on the long-term deformation from sea-level indicators during Marine Isotope Stage (MIS) 5 and the Upper Miocene, as
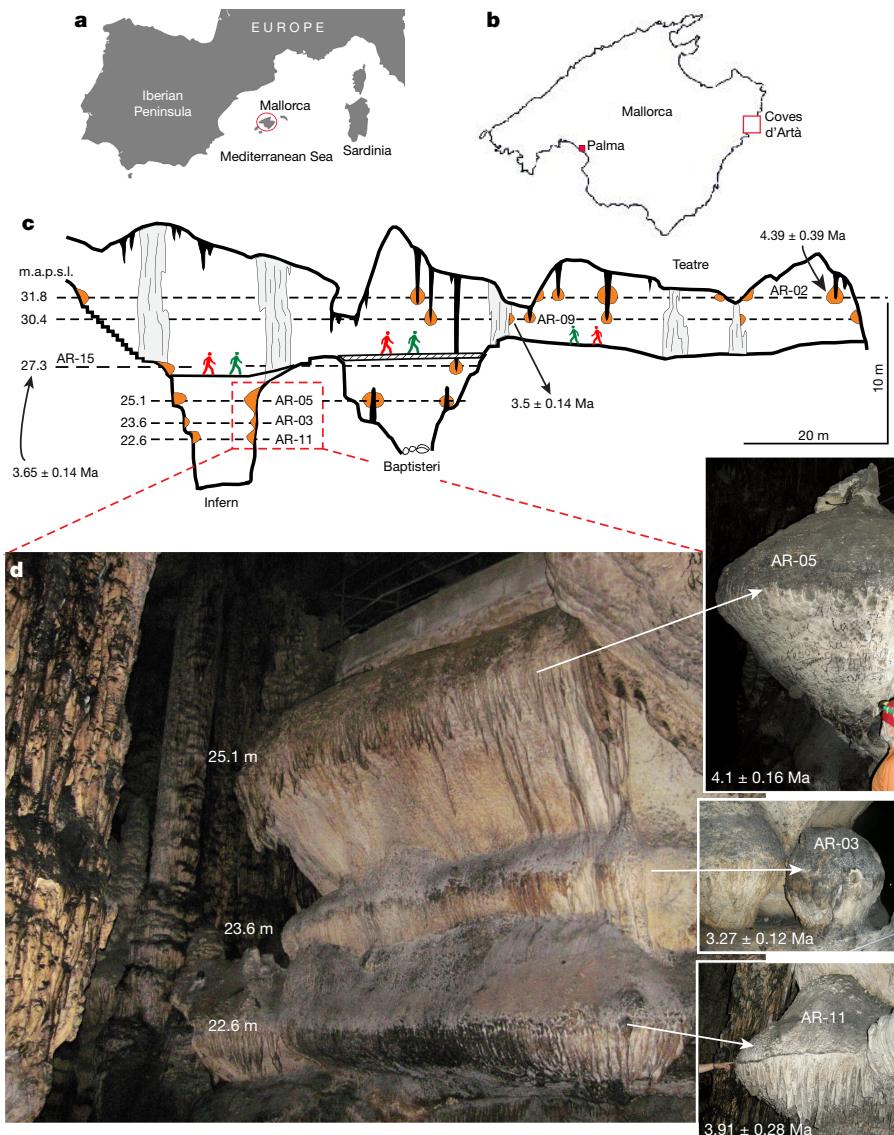
**Fig. 1 | POS in Coves d'Artà. a**, Map showing Mallorca (red circle) in the western Mediterranean. **b**, Location of Coves d'Artà on the island. **c**, Longitudinal profile through the lower section of the cave showing the present-day elevation and ages of the six POS horizons and the sampling sites. **d**, POS at three elevations within the Infern Room with close-up views (insets). Maps (**a**, **b**) are available under CC Public Domain License from https://pixabay.com/illustrations/map-europe-world-earth-continent-2672639/ and https://pixabay.com/illustrations/mallorca-map-land-country-europe-968363/, respectively.

well as a comparison between the relative elevations of the six POS and the GMSL change from three GMSL curves (see Methods). To assess the GMSL we account for uncertainties in the elevation measurement and the indicative range, and correct for GIA, long-term uplift, and thermal expansion (Fig. 2, Table 1; see Methods). Because we correct for thermal expansion, our GMSL estimates throughout the text are sea-level-equivalent ice-volume changes. For better comparison with published estimates, Table 1 additionally includes GMSL estimates without the correction for thermal expansion. Applying all corrections results in a non-Gaussian distribution for our reconstructed GMSL,

of which we report the mode (that is, the most likely value) as our best estimate and in parentheses the 16th and 84th percentiles as our uncertainty range.

Each POS-derived sea-level still stand and the temporally coincident major climatic events evidenced in either marine or terrestrial records are discussed in chronological order and reported in Table 1. The oldest sample (AR-02 at 4.39 ± 0.39 Ma) yields a GMSL estimate of 23.5 m.a.p.s.l. (9.0–26.7 m.a.p.s.l.). Its growth is coincident with an interval considered to be probably the warmest during the Pliocene (~4.4–4.0 Ma, the Pliocene Climatic Optimum), with global mean

**Table 1 | Sample information and results**

| Sample code | Sample elevation (m.a.p.s.l.) | Indicative range (m) | Age (Myr) | Sample mineralogy | Maximum thickness (m) | GIA correction (m) | Uplift correction (m) | Thermal expansion correction (m) | GMSL (m.a.p.s.l.) | GMSL without correction for thermal expansion (m.a.p.s.l.) |
|---|---|---|---|---|---|---|---|---|---|---|
| AR-02 | 31.8 ± 0.25 | 0.55 | 4.39 ± 0.39 | Calcite | 0.20 | 1.3 ± 3.1 | 9.0 (2.5–19.4) | 1.6 ± 0.6 | 23.5 (9.0–26.7) | 25.1 (10.6–28.3) |
| AR-05 | 25.1 ± 0.25 | 1.20 | 4.10 ± 0.16 | Aragonite | 0.80 | 1.5 ± 3.1 | 8.4 (2.3–18.2) | 1.5 ± 0.5 | 16.9 (3.5–20.2) | 18.4 (4.9–21.6) |
| AR-11 | 22.6 ± 0.25 | 0.85 | 3.91 ± 0.28 | Calcite | 0.50 | 1.4 ± 3.1 | 8.0 (2.2–17.3) | 1.4 ± 0.5 | 14.7 (2.0–18.0) | 16.1 (3.4–19.4) |
| AR-15 | 27.3 ± 0.25 | 0.50 | 3.65 ± 0.14 | Aragonite | 0.25 | 1.7 ± 2.9 | 7.4 (2.1–16.2) | 1.3 ± 0.5 | 19.5 (7.6–22.6) | 20.8 (8.9–23.9) |
| AR-09 | 30.4 ± 0.25 | 0.70 | 3.50 ± 0.14 | Aragonite | 0.40 | 1.9 ± 2.9 | 7.1 (2.0–15.5) | 1.2 ± 0.4 | 22.5 (11.3–25.7) | 23.7 (12.5–27.0) |
| AR-03 | 23.6 ± 0.25 | 0.55 | 3.27 ± 0.12 | Aragonite | 0.30 | 1.8 ± 2.7 | 6.7 (1.8–14.5) | 1.2 ± 0.4 | 16.2 (5.6–19.2) | 17.4 (6.8–20.3) |

The age uncertainties are reported as 2σ absolute values. Uplift correction shows the median value and the 16th and 84th percentiles in parentheses as uncertainty bounds. The GMSL shows the mode and 16th and 84th percentiles in parentheses as uncertainty bounds. The GMSL estimates include a correction for GIA, long-term uplift, and thermal expansion. Uncertainties in these corrections are 1σ. All reported corrections are subtracted from the sample elevation to obtain the GMSL. In the last column we provide GMSL estimates that are corrected only for GIA and long-term uplift, but not for thermal expansion.
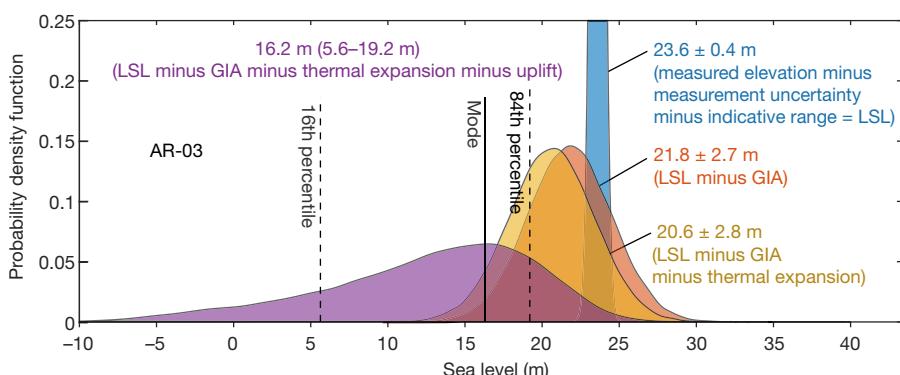
**Fig. 2 | GIA and other corrections.**
Contribution of different corrections (GIA, uplift and thermal expansion) and uncertainties when inferring the GMSL from the POS elevation (this breakdown is for AR-03; see Extended Data Fig. 8 for all POS). Probability density function of the POS elevation with consecutive corrections for the measurement and indicative range leading to an estimate of local sea level (LSL; blue), GIA (orange), long-term uplift (purple curve) and thermal expansion (yellow). We choose the mode (solid black line) as the best estimate and the 16th and 84th percentiles (dashed black line) as the uncertainty range.

temperatures roughly 4 °C higher than pre-industrial values[6] and elevated $CO_2$ (ref. [22]) (Fig. 3a).

A GMSL estimate of 16.9 m.a.p.s.l. (3.5–20.2 m.a.p.s.l.) at 4.1 ± 0.16 Ma is based on a core extracted from AR-05, the thickest POS horizon (top inset in Fig. 1d). A pronounced expansion of the ice sheets (MIS Gi22/Gi20) and the associated sea-level drop was documented at ~4 Ma in several locations in the Northern Hemisphere[23] and in Prydz Bay, Antarctica[24]. Therefore, AR-05 is interpreted to reflect a rather long, muted sea-level still stand that occurred before this cooler Pliocene interval.

Similarly, the inner part of AR-11 (bottom inset in Fig. 1d), documents a lower GMSL of 14.7 m.a.p.s.l. (2.0–18.0 m.a.p.s.l.) at 3.91 ± 0.28 Ma. Sample AR-15 marks a GMSL of 19.5 m.a.p.s.l. (7.6–22.6 m.a.p.s.l.) at 3.65 ± 0.14 Ma and overlaps with the Northern Hemisphere glacial interval known as MIS MG 12 (~3.7–3.6 Ma), which is considered to represent the Early/Late Pliocene transition[10,23,25]. Terrestrial and marine data indicate that the Northern Hemisphere glaciation onset was around 3.6 Ma (refs [23,25]), but other records from the Northern and Southern hemispheres suggest that relatively warm climatic conditions prevailed until 3.5 Ma (ref. [26]). This observation is supported

by the presence of yet another horizon formed in Coves d'Artà (AR-09) at 3.5 ± 0.14 Ma that indicates a GMSL of 22.5 m.a.p.s.l. (11.3–25.7 m.a.p.s.l.).

Sample AR-03, with an inferred GMSL of 16.2 m.a.p.s.l. (5.6–19.2 m.a.p.s.l.) documents the youngest prominent Pliocene horizon in Coves d'Artà with an age of 3.27 ± 0.12 Ma (middle inset in Fig. 1d), which probably formed at the onset of the MPWP. If one accounts for GIA but assumes no long-term deformation, the GMSL during the MPWP is predicted to be 20.6 ± 2.8 m (Fig. 2). Overall, our calculations are most consistent with previous sea-level estimates that are based on benthic foraminifera[10–12] (Fig. 3b). Our results are noticeably lower than the GMSL based on data by Rohling et al.[2] and higher than those based on ice-sheet modelling[4] (Fig. 3b). The POS-based inferred GMSL for the MPWP is at the lower end of some previous estimates (25 ± 5 m and 22 ± 10 m)[11,17] and overlaps with others (9–13.5 m)[18].

The inherent uncertainties in predicting sea-level rise, when warming is triggered by future anthropogenic emissions of greenhouse gases, emphasize the importance of a better constraint on ice-sheet sensitivity to prolonged warming[27]. The present-day East Antarctic Ice Sheet is less vulnerable to warming than the Greenland Ice Sheet
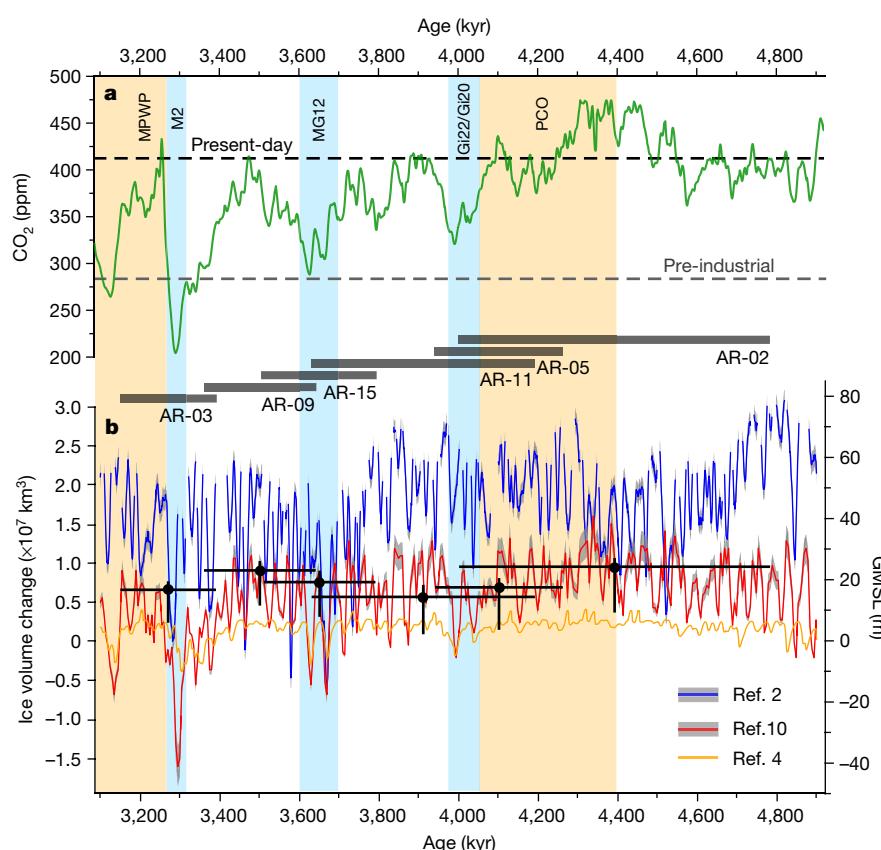


**Fig. 3 | Pliocene sea-level and $CO_2$ concentration changes. a**, Model-based $CO_2$ reconstruction[21] and relevant warm (orange bands) and cold (blue bands) climatic periods. **b**, Inferred GMSL and ice volume from Mallorcan POS are shown as black markers (age uncertainties are 2σ; the GMSL of the marker corresponds to the mode; uncertainties are 16th and 84th percentiles). The sample code for each POS is indicated on the grey band between panels. Coloured curves show three different GMSL reconstructions (uncertainties on GMSL curves are 1σ). See Methods for the derivation of the GMSL curves. PCO, Pliocene Climatic Optimum.

and the West Antarctic Ice Sheet[28]. Considering this, the equivalent ice volume required for an MPWP GMSL of 16.2 m.a.p.s.l. (5.6–19.2 m.a.p.s.l.) (AR-03) would indicate a near or full collapse of the modern Greenland Ice Sheet (7.4 m GMSL rise)[29] and vulnerable sectors of the West Antarctic Ice Sheet (3.3 m GMSL rise)[30], plus a contribution of up to 5.5 m from more stable sectors of the West Antarctic Ice Sheet and from the East Antarctic Ice Sheet. Given that the marine sectors of the East Antarctic Ice Sheet hold an ice volume equivalent to 19 m of GMSL rise[31], our estimates do not require a contribution from land-based sectors of the East Antarctic Ice Sheet. Instead, they indicate a possible retreat of East Antarctic Ice Sheet in some marine-based sectors, but stability of land-based ice masses, which is consistent with proxies from Antarctica[26,32] and ice-sheet models[7]; we are thus able to narrow in on closing the sea-level budget for the MPWP.

Given that global average temperatures during the MPWP were 2–3 °C higher than pre-industrial values[5] and $CO_2$ concentration was 400 ppm (ref. [9]), our results indicate that an ice volume equivalent to a GMSL change of 16.2 m.a.p.s.l. (5.6–19.2 m.a.p.s.l.) may eventually melt (over hundreds to thousands of years) if future temperatures stabilize at that level of warming. Given present-day melt patterns[26], this sea-level rise is likely to be sourced from a collapse of both Greenland and the West Antarctic ice sheets. A temperature increase to 4 °C above pre-industrial levels is comparable to conditions during the Pliocene Climatic Optimum[6] with a GMSL estimate of 23.5 m.a.p.s.l. (9.0–26.7 m.a.p.s.l.), which indicates further ice melt if temperatures stabilize at this higher level. Thermal expansion is expected to cause additional sea-level rise in these scenarios.

Projections of future sea-level rise that are tuned to fit the GMSL during the MPWP suggest that the Antarctic contribution to sea level by 2100 will be either $1.05 \pm 0.30$ m or $0.64 \pm 0.49$ m, if its GMSL contribution during the MPWP was assumed to be 10–20 m.a.p.s.l. or 5–15 m.a.p.s.l., respectively (under scenario RCP8.5)[7]. However, Edwards[16] questioned the way marine ice cliff instability and hydrofracturing is parameterized in this work and found the latter interval to be more probable. Our AR-03-derived MPWP sea-level range (5.6–19.2 m.a.p.s.l.), which includes contributions from all ice sheets, is also more in line with the lower end of the estimates of DeConto and Pollard[7]. Nonetheless, this highlights (1) the need for further work to reduce uncertainty in MPWP GMSL estimates, and (2) the importance of our results for the even warmer Pliocene Climatic Optimum (9.0–26.7 m.a.p.s.l.). These data will serve as critical inputs for future climate model development and calibration that will improve confidence in sea-level projections. Hence, deciphering the GMSL during Pliocene warmth is critical for our ability to forecast, adapt to, and lessen the effect of future global warming on humanity.

## Online content

1. Dutton, A. et al. Sea-level rise due to polar ice-sheet mass loss during past warm periods. *Science* **349**, aaa4019 (2015).
2. Rohling, E. J. et al. Sea-level and deep-sea-temperature variability over the past 5.3 million years. *Nature* **508**, 477–482 (2014); corrigendum **510**, 432 (2014).
3. Raymo, M. E., Kozdon, R., Evans, D., Lisiecki, L. & Ford, H. L. The accuracy of mid-Pliocene $\delta^{18}O$-based ice volume and sea level reconstructions. *Earth Sci. Rev.* **177**, 291–302 (2018).
4. De Boer, B., van de Wal, R. S. W., Bintanja, R., Lourens, L. J. & Tuenter, E. Cenozoic global ice-volume and temperature simulations with 1-D ice-sheet models forced by benthic $\delta^{18}O$ records. *Ann. Glaciol.* **51**, 23–33 (2010).
5. Haywood, A. M. et al. Large-scale features of Pliocene climate: results from the Pliocene Model Intercomparison Project. *Clim. Past* **9**, 191–209 (2013).
6. Fedorov, A. V. et al. Patterns and mechanisms of early Pliocene warmth. *Nature* **496**, 43–49 (2013).
7. DeConto, R. M. & Pollard, D. Contribution of Antarctica to past and future sea-level rise. *Nature* **531**, 591–597 (2016).
8. Zachos, J. C., Shackleton, N. J., Revenaugh, J. S., Palike, H. & Flower, B. P. Climate response to orbital forcing across the Oligocene-Miocene boundary. *Science* **292**, 274–278 (2001).
9. Pagani, M., Liu, Z. H., LaRiviere, J. & Ravelo, A. C. High Earth-system climate sensitivity determined from Pliocene carbon dioxide concentrations. *Nat. Geosci.* **3**, 27–30 (2010).
10. Lisiecki, L. E. & Raymo, M. E. A Pliocene-Pleistocene stack of 57 globally distributed benthic $\delta^{18}O$ records. *Paleoceanography* **20**, PA1003 (2005).
11. Dwyer, G. S. & Chandler, M. A. Mid-Pliocene sea level and continental ice volume based on coupled benthic Mg/Ca palaeotemperatures and oxygen isotopes. *Phil. Trans. R. Soc. A* **367**, 157–168 (2009).
12. Naish, T. R. & Wilson, G. S. Constraints on the amplitude of Mid-Pliocene (3.6–2.4 Ma) eustatic sea-level fluctuations from the New Zealand shallow-marine sediment record. *Phil. Trans. R. Soc. A* **367**, 169–187 (2009).
13. Rovere, A. et al. The Mid-Pliocene sea-level conundrum: glacial isostasy, eustasy and dynamic topography. *Earth Planet. Sci. Lett.* **387**, 27–33 (2014).
14. Raymo, M. E., Mitrovica, J. X., O'Leary, M. J., DeConto, R. M. & Hearty, P. J. Departures from eustasy in Pliocene sea-level records. *Nat. Geosci.* **4**, 328–332 (2011).
15. Rowley, D. B. et al. Dynamic topography change of the Eastern United States since 3 million years ago. *Science* **340**, 1560–1563 (2013).
16. Edwards, T. L. et al. Revisiting Antarctic ice loss due to marine ice-cliff instability. *Nature* **566**, 58–64 (2019).
17. Miller, K. G. et al. High tide of the warm Pliocene: implications of global sea level for Antarctic deglaciation. *Geology* **40**, 407–410 (2012).
18. Winnick, M. J. & Caves, J. K. Oxygen isotope mass-balance constraints on Pliocene sea level and East Antarctic Ice Sheet stability. *Geology* **43**, 879–882 (2015).
19. Polyak, V. J. et al. A highly resolved record of relative sea level in the western Mediterranean Sea during the last interglacial period. *Nat. Geosci.* **11**, 860–864 (2018).
20. Kendall, R. A., Mitrovica, J. X. & Milne, G. A. On post-glacial sea level—II. Numerical formulation and comparative results on spherically symmetric models. *Geophys. J. Int.* **161**, 679–706 (2005).
21. Moucha, R. & Ruetenik, G. A. Interplay between dynamic topography and flexure along the U.S. Atlantic passive margin: insights from landscape evolution modeling. *Global Planet. Change* **149**, 72–78 (2017).
22. Stap, L. B. et al. $CO_2$ over the past 5 million years: continuous simulation and new $\delta^{11}B$-based proxy data. *Earth Planet. Sci. Lett.* **439**, 1–10 (2016).
23. De Schepper, S., Gibbard, P. L., Salzmann, U. & Ehlers, J. A global synthesis of the marine and terrestrial evidence for glaciation during the Pliocene Epoch. *Earth Sci. Rev.* **135**, 83–102 (2014).
24. Passchier, S. Linkages between East Antarctic Ice Sheet extent and Southern Ocean temperatures based on a Pliocene high-resolution record of ice-rafted debris off Prydz Bay, East Antarctica. *Palaeocanogr. Paleoclimatol.* **26**, PA4204 (2011).
25. Mudelsee, M. & Raymo, M. E. Slow dynamics of the Northern Hemisphere glaciation. *Paleoceanography* **20**, PA4022 (2005).
26. Williams, T. et al. Evidence for iceberg armadas from East Antarctica in the Southern Ocean during the late Miocene and early Pliocene. *Earth Planet. Sci. Lett.* **290**, 351–361 (2010).
27. Foster, G. L. & Rohling, E. J. Relationship between sea level and climate forcing by $CO_2$ on geological timescales. *Proc. Natl Acad. Sci. USA* **110**, 1209–1214 (2013).
28. Church, J. A. et al. Sea level change. *Climate Change 2013: The Physical Science Basis* (Cambridge Univ. Press, 2013).
29. Morlighem, M. et al. BedMachine v3: complete bed topography and ocean bathymetry mapping of Greenland from multibeam echo sounding combined with mass conservation. *Geophys. Res. Lett.* **44**, 11051–11061 (2017).
30. Bamber, J. L., Riva, R. E., Vermeersen, B. L. & LeBrocq, A. M. Reassessment of the potential sea-level rise from a collapse of the West Antarctic Ice Sheet. *Science* **324**, 901–903 (2009).
31. Fretwell, P. et al. Bedmap2: improved ice bed, surface and thickness datasets for Antarctica. *Cryosphere* **7**, 375–393 (2013).
32. Shakun, J. D. et al. Minimal East Antarctic Ice Sheet retreat onto land during the past eight million years. *Nature* **558**, 284–287 (2018).

## METHODS

**POS as precise sea-level index points.** POS are meaningful sea-level index points because they provide spatial geographic positioning, accurate elevation and absolute ages. In addition, the growth of POS has a clear indicative meaning, which includes the 'indicative range', that is, the elevation range over which a sea-level indicator forms, and the 'reference water level', meaning the distance between the sea-level indicator and sea level[33]. The indicative range for POS is the vertical extent over which the carbonate encrustations occur and the reference water level is zero at the widest part of the POS (Extended Data Fig. 1). This is true because the widest part (maximum thickness) of the POS that formed on cave walls or other speleothems that are in continuous contact with the fluctuating water table corresponds to mean sea level, whereas the vertical spread constrains the tidal range (top and bottom insets in Fig. 1d). The shape of POS formed on pre-existing vadose (air-filled caves) speleothems depends on their size and morphology and for how long they were immersed in the cave's brackish water (Extended Data Fig. 1). If, for example, only the tip of the stalactites become submerged, the resulting POS will be an asymmetric knob-like carbonate encrustation with its thickest part denoting the mean sea level and narrowing upward to the highest tide range (see the asterisked POS in Extended Data Fig. 1). Thus, the most optimal POS form when the speleothem is immersed throughout the full vertical tidal range, producing an oval (sometimes spherical) or fusiform-shaped POS (middle inset in Fig. 1d, Extended Data Fig. 1a). Except for AR-02, which is a knob-like POS, all the others were either standalone (Extended Data Fig. 1a) or cave-wall (Extended Data Fig. 1b) structures.

Prior to our sampling campaign, a detailed topographic survey was conducted starting from a reference point outside the cave and situated at the present sea level, using a SUUNTO optical clinometer and a BOSCH DLE 50 Professional laser distance meter. The errors associated with measuring the elevation of the POS levels relative to the mean tidal level (after correcting for the barometric pressure) are less than 0.25 m. Owing to the large size of some POS horizons, cores were drilled using a commercial cordless hand-held Hilti rotary hammer drill.

**U–Pb analytical methods.** Absolute isotope-dilution U–Pb ages were measured using a Thermo Neptune multi-collector inductively coupled mass spectrometer. The analytical methodology is reported in Decker et al.[34]. All sub-samples used for U–Pb measurements were clean, well crystallized aragonite or calcite pieces. A calibrated mixed $^{229}$Th–$^{233}$U–$^{236}$U–$^{205}$Pb spike was used. The column chemistry for all of these analyses uses $1 \times 8$ chloride form 200–400 mesh anion resin. Each element isotope system was run separately. Pb runs were measured as standard-sample-sample-standard runs using the Pb standard NBS-981. $^{234}$U signals too small to measure in the centre Faraday cup were measured using the secondary electron multiplier with the gain between the secondary electron multiplier and Faraday cups established using the U standard NBL-112. $^{230}$Th and $^{232}$Th were also measured using the secondary electron multiplier and a Faraday cup, respectively, and a Th in-house standard. $^{230}$Th/$^{238}$U was measured to check samples for U and Th isotope secular equilibrium.

**Age calculations.** Reduction of data was performed using PBDAT[35] and ISOPLOT[36]. Our measured procedural blanks are 5 pg of $^{232}$Th, <0.1 fg of $^{230}$Th, 20 pg of $^{238}$U, and ~15 pg of $^{208}$Pb. Decay constants for $^{238}$U, $^{235}$U, $^{234}$U and $^{230}$Th are from Schoene et al.[37] and Cheng et al.[38]. The isotopic values for the NBS-981 Pb standard reported by the National Institute of Standards were used in the sample-standard method[39]. U–Pb ages were corrected for initial disequilibrium assuming an initial $^{234}$U/$^{238}$U activity of 1.75, based on the average initial $\delta^{234}$U of over 200 samples from the same island, ranging from Holocene to Pleistocene, dated using the U–Th method. Additionally, results of $\delta^{234}$U measurements of the present-day brackish water in which POS precipitate (as described earlier) indicate similar values[19].

Except for AR-02, for which the $^{235}$U–$^{207}$Pb two-dimensional isochron age was more favourable than the three-dimensional isochron ages, all other sample ages were calculated using the U–Pb Concordia-constrained linear three-dimensional isochron, which contains the most complete information on the concordance between the two decay schemes and common Pb (see Extended Data Fig. 2).

In certain plots, a limited number of discrepant data have been excluded from the fit, for reasons which are not under statistical control, such as non-uniform Pb composition and possibly variable initial $\delta^{234}$U. We excluded the subsamples that deviated markedly from the isochron lines in order to make sure that the isotopic analysis did not mix growth zones of domains with different ages. Considering that our samples precipitated in a phreatic environment, growth layers are hardly noticeable, so sampling along a single layer was challenging. Nevertheless, we stress that our analysed datasets are large enough to allow us to distinguish any discordant data, and as a result, the calculated ages are considered to be accurate within the given uncertainties. To have a superior control on the random uncertainties and for a better recognition of the outliers, we analysed between 9 and 15 subsamples for each isochron to obtain the ages, their uncertainties, and associated mean-square-weighted deviation values.

**GIA modelling.** For the GIA correction we adopt a one-dimensional, self-gravitating, Maxwell viscoelastic Earth model[20], which accounts for shoreline migration and the feedback into Earth's rotation axis. Following Raymo et al.[14] we separate the GIA correction into two contributions: (1) the amount of residual deformation that is due to the most recent Pleistocene glacial cycles, and (2) the amount of deformation due to the smaller Pliocene ice age cycles. For both corrections we use a set of different Earth structure profiles that vary in lithospheric thickness (48 km, 71 km and 96 km), upper-mantle viscosity ($3 \times 10^{20}$ Pa s, $5 \times 10^{20}$ Pa s), and lower-mantle viscosity ($3 \times 10^{21}$ Pa s, $5 \times 10^{21}$ Pa s, $7 \times 10^{21}$ Pa s, $10 \times 10^{21}$ Pa s, $20 \times 10^{21}$ Pa s and $30 \times 10^{21}$ Pa s), producing 36 different radial viscosity profiles. For the elastic and density structure we assume the seismic reference model PREM[40]. In this approach we neglect explicit three-dimensional variations in viscosity; although these unarguably exist, their pattern and magnitude are poorly constrained. Owing to these uncertainties and the high computational cost associated with these runs we follow the common approach to explore uncertainties in viscosity through a variety of one-dimensional viscosity profiles. We believe that this approach is particularly appropriate because we are not investigating spatial patterns in GIA (for which three-dimensional variability might be important), but focus on only one location.

For the two GIA contributors mentioned above we proceed as follows. (1) We adopt the ice history over the past 3 Myr as described in Raymo et al.[14]. The models indicate a positive GIA correction for most of the Mediterranean (Extended Data Fig. 3a), mainly due to ongoing peripheral bulge collapse associated with the former Fennoscandian Ice Sheet. This sea-level rise implies that Coves d'Artà is currently above its equilibrium sea level and that sea-level indicators that formed during the more equilibrated Pliocene need to be corrected downward. Using one possible viscoelastic Earth model (Extended Data Fig. 3a) indicates that the remaining adjustment at Coves d'Artà is 3.4 m. Employing all 36 model runs yield a mean and standard deviation of $4.5 \pm 2.1$ m for this location (Extended Data Fig. 3).

(2) We constructed new ice models by scaling the height of present-day ice sheets to reproduce a given ice-volume curve. We set up three different ice models based on the LR04 benthic stack[10], the local sea-level reconstructions by Rohling et al.[2], and the ice-sheet model by de Boer et al.[4]. Uncertainties will be assessed by considering these three ice models rather than propagating uncertainties in each of them. However, we acknowledge that large uncertainties exist for each approach and these will be considered in the long-term deformation component (see Methods section 'Estimating long-term deformation'). The ice-volume curves are constructed for each model as follows.

(1) To scale the oxygen isotope signal into ice-volume changes, we assume that 75% of the signal is driven by ice volume (the remaining 25% is driven by temperature variations). This value is in line with Pleistocene ocean temperature estimates obtained using Mg/Ca ratio[17,41]. We further assume a scaling of 0.011‰ per metre of GMSL rise[3,12].

(2) Rohling et al.[2] used planktonic foraminifera and the marginal basin residence time method for the Mediterranean to produce a relative sea-level record for the Strait of Gibraltar (RSLGib$_{observed}$). They further provide a scaling to calculate ice volume (GMSL$_{scaled}$ = $1.23 \times$ RSLGib$_{observed}$ + 0.5), which is based on simulations for two glacial cycles. We use this scaling as a first estimate for ice-volume changes. We next run the sea-level model to calculate local sea-level changes at the Strait of Gibraltar for a variety of viscosity models. We take the mean of these local sea-level estimates (RSLGib$_{calculated}$) to calculate a GIA correction (GIA = RSLGib$_{calculated}$ − GMSL$_{scaled}$). Last, we use this correction to recalculate the GMSL estimate (GMSL = RSLGib$_{observed}$ − GIA). We note that the original sea-level reconstruction has data gaps associated with the African monsoon. We use the interpolated reconstruction here, but exclude data during these gaps in our final comparison (Fig. 3).

(3) De Boer et al.[4] used a set of one-dimensional ice-sheet simulations that are forced by a benthic oxygen isotope record through surface-air temperatures. They separately model five ice sheets (Greenland Ice Sheet, Laurentide, Fennoscandian, West Antarctic Ice Sheet and East Antarctic Ice Sheet) and provide an ice-volume reconstruction over the Cenozoic.

For times during which ice volume was lower than today, we first decrease the height of the Greenland Ice Sheet and West Antarctic Ice Sheet until they are fully collapsed before we start decreasing the height of the East Antarctic Ice Sheet. For ice volumes higher than today we uniformly increase the size of all ice sheets. Ice-rafted debris and other evidence from the Greenland continental shelf indicates that an intermittent ice sheet existed on Greenland during the Pliocene, but that the main expansion probably happened around 3 Ma (ref. [42]). This is largely consistent with our ice reconstructions based on the GMSL scenarios described above. We further tested a scenario in which only the Antarctic Ice Sheet varied over the model run and the Greenland Ice Sheet was not present. This introduced only minor differences (<0.6 m) in the GIA correction at Coves d'Artà and is therefore not considered further here.

Calculations were done with a temporal resolution of 1,000 years and a spatial resolution up to spherical harmonic degree 256. The GMSL throughout the paper is directly proportional to total ice volume (not only ice above flotation) given by the equation:

$$\text{GMSL} = \frac{\text{Ice volume changes} \times \rho_{\text{ice}}}{\text{Percentage oceanic area} \times \rho_{\text{water}}}$$

For the percentage oceanic area we use a fixed value of 71.1% and further use a water density of 1,000 kg m$^{-3}$ and an ice density of 920 kg m$^{-3}$.

The resulting GMSL and local sea-level curve for Coves d'Artà for one run is shown in the Extended Data Fig. 4a along with a snapshot of sea level at 3.244 Ma (Extended Data Fig. 4b). Local sea level (coloured curves) is lower than the GMSL (black curves) during times when ice sheets were collapsed. This difference is driven by the accommodation space that is created by the collapsed marine-based sectors. Water flowing back to these areas causes high rates of sea-level rise (for example, in West Antarctica), which is averaged out by an overall drawdown in sea level in the far field (Extended Data Fig. 4a). These runs are relative to the start of the model time (outside the range shown in Extended Data Fig. 4b), which was set to be the present-day ice configuration.

The full GIA correction is given by a combination of both effects described above. The resulting local sea-level prediction is shown in Extended Data Fig. 5. Our calculations show that local sea level tracks the GMSL relatively closely when the GMSL is high (Extended Data Fig. 5a) because the two contributions discussed above cancel out. When the GMSL is low (Extended Data Fig. 5c) the GIA signal is mainly driven by the ongoing adjustment to the Last Glacial Maximum, which means the GIA correction is positive (Extended Data Fig. 3; Extended Data Fig. 5f).

The GIA correction is given by the difference between the local sea level at Coves d'Artà and the GMSL (GIA = RSLArtà$_{\text{calculated}}$ − GMSL). To quantify the GIA correction and its uncertainty for each POS we take a temporal average and standard deviation of all GIA models (using all three GMSL curves) over the time period of the sea-level indicators. We consider only corrections during intermediate and warm periods, assuming that these provide more favourable conditions for POS to form. This assumption is based on the fact that POS form during sea-level still stands and the Pleistocene still stands occur more frequently during intermediate and warm periods than during glacials[43]. Weakening this assumption would lead to a slightly larger uncertainty in the GIA correction. To include this assumption, we identify the GMSL values (and their associated GIA correction) that fall within the age range of each POS, and average over the GIA correction that corresponds to the highest 50% of GMSL values (80% in the case of AR-03 to avoid a bias towards MIS M2, Extended Data Fig. 5). We combine values from the three ice models to calculate the mean and standard deviation in the GIA correction for each data point (Table 1).

**Estimating long-term deformation.** Mallorca is generally described as stable with very little to no long-term deformation[44,45]. However, even a small amount of deformation (uplift or subsidence) can substantially affect our results. We therefore investigate constraints on the long-term deformation of the island based on sea-level records from other time periods. We further attempt to quantify uncertainties on the possible amount of uplift by comparing the relative elevations of the six POS to the relative GMSL change in the three curves described above[2,12].

POS in Mallorca that date to MIS 5a are above present sea level today[46]. The GIA correction for this location is small and the GMSL was possibly around present levels but probably lower[47], indicating potential uplift. POS dated to the last interglacial are found at 2.15 ± 0.75 m.a.p.s.l.[19]. Given the uncertainties in the eustatic estimate during the last interglacial and the GIA correction, it is difficult to identify uplift or subsidence. Late Miocene reefs that crop out at 65 m above present at Cap Blanc[48] and up to 70 m in the hinterland of Mallorca[44,49] are also high compared to global average values (once corrected for GIA), but Late Miocene GMSL estimates are much more uncertain[50]. Given that local sea-level estimates tend to be high, it is unlikely that subsidence occurred at Coves d'Artà. We will therefore assume that the record presented here is not affected by subsidence; however, the evidence reported above does not exclude slight long-term uplift.

We estimate the amount of possible long-term uplift based on relative sea-level changes across the POS record and its comparison to the three GMSL reconstructions. Unlike our GIA calculation, uncertainties in the different GMSL curves will be important for this analysis and we therefore choose the following approach to quantify the respective uncertainties.

(1) For the GMSL curve that is obtained from the LR04 benthic record[10] we widen the range for scaling oxygen isotope values into GMSL change to 0.008 and 0.011‰ per metre[3,17] to produce two end-member GMSL curves. We assume that the mean of these two curves is the most likely estimate, and the two end-member curves constitute 1σ uncertainty. This results in a wider uncertainty than if we only assumed that the two end-members would span the range of possible GMSL curves. This extended uncertainty is meant to implicitly include further

uncertainties associated with estimating the amount of oxygen isotope signal that is driven by ice volume versus temperature.

(2) The GMSL curve based on the data by Rohling et al.[2] is obtained through the equation GMSL = RSLGib$_{\text{observed}}$ − GIA, excluding the gaps in their planktonic foraminiferal dataset due to maxima in the African monsoon. We calculate the uncertainty at each timestep as the root mean square of the uncertainty associated with the relative sea-level observation (provided by Rohling et al.[2]) and the uncertainty associated with the GIA correction for the Strait of Gibraltar that is caused by varying viscosity models.

(3) De Boer et al.[4] do not quantify an uncertainty in their estimate, so we do not show it in Fig. 3. However, for our long-term deformation analysis we do want to attempt an uncertainty estimate. De Boer et al.[4] do two sensitivity tests in which they vary the deep-water to surface-air temperature coupling, and the temperature difference for the Northern Hemisphere ice sheets. They found that their predicted ice volume was relatively insensitive to these variations with a largest difference between runs of ~3.5 m during the Pliocene. Here we assume that their best-fit curve represents a mean estimate and that the 1σ uncertainty is 1 m.

While the GMSL reconstructions described above and shown in Fig. 3 vary widely, relative changes, for example, the change in GMSL from the Pliocene Climatic Optimum to the MPWP could be more robust. We quantify bounds on the amount of sea-level change that occurred between the different POS from our three GMSL reconstructions. For each reconstruction we generate 500 possible GMSL curves, sampling the uncertainty in each of them (grey bands in Extended Data Fig. 6a–c are 1σ). We next bin the GMSL values that fall within the age range of each POS and identify values that are within a certain (average to high) percentile range to reflect intermediate and warm periods. For the purpose of calculating an uplift rate we again assume that it is more likely that these are the periods during which our POS formed. We vary the lower bound between the 40th, 50th and 60th percentiles and the upper bound between the 90th, 95th and 99th percentiles. Extended Data Fig. 6a–c shows the range of GMSL values considered for each POS as grey boxes for a scenario of the 50th percentile lower bound and 99th percentile upper bound. We do a Monte Carlo simulation in which we randomly sample 'synthetic' sea-level indicator elevations from the respective GMSL ranges, that is, we pick one random point from each grey box in Extended Data Fig. 6a–c. We next calculate the change in sea level in these synthetic data relative to the youngest data point. These changes are compared to the observed changes in sea level (GIA corrected) to which we add a constant uplift rate. We choose this uplift rate to be constant, but vary its magnitude in each iteration. This assumes that long-term deformation is linear to first order over the Plio-Pleistocene, which is supported by studies of mantle convection that show that uplift rates related to dynamic topography are relatively constant over a few million years[51]. We consider the synthetic data to be a good fit to the observed data if their difference does not exceed 3 m for a given data point. This value is chosen because it is the root mean square of the average GIA uncertainty, the measurement uncertainty, and half the indicative range. We record the uplift rates that are successful, that is, produce a good fit.

Extended Data Fig. 6d–f shows histograms of these successful uplift rates for a scenario of 50th percentile lower bound and 99th percentile upper bound. GMSL curves that are based on Rohling et al.[2] and the LR04 benthic record[10] favour small uplift rates, because the variability within these curves is already large enough to represent the variability within the POS elevation data. However, there is a tail towards higher uplift rates. The GMSL curve based on de Boer et al.[4] is notably flatter and, in order to match the elevation variability in the POS results, these data require a modest amount of uplift. We produce a joint distribution, in which we combine all successful uplift rates (Extended Data Fig. 6g). We repeat this procedure varying the lower-bound and upper-bound percentiles as described above to obtain nine joint distributions (Extended Data Fig. 7a–i). Last, we combine the successful uplift rates from all nine joint distributions to obtain our final distribution, which is the basis for our long-term uplift correction (Extended Data Fig. 7j). The median uplift rate we obtain is 2.0 m Myr$^{-1}$ (0.6–4.4 m Myr$^{-1}$; uncertainties constitute the 16th and 84th percentiles). The most likely uplift rate (highest number of successful runs) is 0 m Myr$^{-1}$. These rates are within the uncertainty range of uplift estimates for Mallorca based on the MIS 5e sea-level estimate[52].

**Correcting local sea level to obtain the GMSL.** We produce a probability density function (PDF) for the elevation of each POS. We first take into account the uncertainty of the measurement (0.25 m) and the indicative meaning (half the indicative range). Since we assume that these errors are normally distributed, the resulting PDF (blue curve, Fig. 2c) is also normally distributed. We next correct for GIA assuming the values shown in Extended Data Fig. 5, which results in the red curve (Fig. 2c). In the next step, we account for thermal expansion. We assume a linearly increasing effect of thermal expansion, 0.39 m of GMSL rise per degree Celsius[53] and 4 °C of warming at the beginning of the Pliocene Climatic Optimum (4.4–4.0 Ma)[6]. To calculate the thermal expansion correction from this rate we require the age of each POS. We randomly sample the age of the POS from its uncertainty range. This correction is again normally distributed, resulting in a new PDF that is

normally distributed as well (yellow curve, Fig. 2c). Last, we account for long-term deformation. We use the distribution that we obtain as described above (Extended Data Fig. 7j) for this correction. We sample the age of the POS again to translate an uplift rate into the amount of total uplift. The resulting PDF (purple curve, Fig. 2c) is off-centred owing to the skewness in the long-term deformation distribution. We therefore do not quantify uncertainties as standard deviations but instead determine the mode (most likely value) as our preferred value (GMSL) and the 16th and 84th percentiles as error bounds. We use a kernel with 1-m bandwidth to calculate the mode. Extended Data Fig. 8 shows the PDFs for the GMSL estimate for each POS after all corrections have been applied. Extended Data Table 1 includes additional percentiles of the GMSL estimate (10th, 33th, 50th, 66th and 90th) in line with IPCC's likelihood scale.

## Data availability

The data produced in this study are available at the NOAA (https://www.ncdc.noaa.gov/paleo/study/27530) and Pangaea (https://doi.pangaea.de/10.1594/PANGAEA.905851) data repositories.

## Code availability

The computer code used to do the sea-level (GIA) calculation, written in MATLAB, is available on github: https://github.com/jaustermann/SLcode.

33. Rovere, A. et al. The analysis of Last Interglacial (MIS 5e) relative sea-level indicators: reconstructing sea-level in a warmer world. *Earth Sci. Rev.* **159**, 404–427 (2016).
34. Decker, D. D. et al. U–Pb dating of cave spar: a new shallow crust landscape evolution tool. *Tectonics* **37**, 208–223 (2018).
35. Ludwig, K. R. & Titterington, D. M. Calculation of [230]Th/U isochrons, ages, and errors. *Geochim. Cosmochim. Acta* **58**, 5031–5042 (1994).
36. Ludwig, K. R. *User's Manual for Isoplot 3.75: A Geochronological Toolkit for Microsoft Excel* http://www.bgc.org/isoplot_etc/isoplot/isoplot3_75-4_15manual.pdf (Berkeley Geochronology Center Special Publication 5, 2012).
37. Schoene, B. et al. Reassessing the uranium decay constants for geochronology using ID-TIMS U–Pb data. *Geochim. Cosmochim. Acta* **70**, 426–445 (2006).
38. Cheng, H. et al. Improvements in [230]Th dating, [230]Th and [234]U half-life values, and U–Th isotopic measurements by multi-collector coupled plasma mass spectrometry. *Earth Planet. Sci. Lett.* **371–372**, 82–91 (2013).
39. Yuan, H. et al. Evaluation of lead isotope compositions of NIST NBS 981 measured by thermal ionization mass spectrometer and multiple-collector inductively coupled plasma mass spectrometer. *Solid Earth Sci.* **1**, 74–78 (2016).
40. Dziewonski, A. M. & Anderson, D. L. Preliminary reference Earth model. *Phys. Earth Planet. Inter.* **25**, 297–356 (1981).
41. Elderfield, H. et al. Evolution of ocean temperature and ice volume through the Mid-Pleistocene Climate Transition. *Science* **337**, 704–709 (2012).
42. Bierman, P. R., Shakun, J. D., Corbett, L. B., Zimmerman, S. R. & Rood, D. H. A persistent and dynamic East Greenland Ice Sheet over the past 7.5 million years. *Nature* **540**, 256–260 (2016).
43. Shakun, J. D., Lea, D. W., Lisiecki, L. E. & Raymo, M. E. An 800-kyr record of global surface ocean $\delta^{18}$O and implications for ice volume-temperature coupling. *Earth Planet. Sci. Lett.* **426**, 58–68 (2015).
44. Pomar, L. & Ward, W. C. in *Sequence Stratigraphy and Depositional Response to Eustatic, Tectonic and Climatic Forcing* (ed. Haq, B. U.) Ch. 4, 87–112 (Kluwer, 1995).
45. Just, J., Hübscher, C., Betzler, C., Lüdmann, T. & Reicherter, K. Erosion of continental margins in the Western Mediterranean due to sea-level stagnancy during the Messinian Salinity Crisis. *Geo-Mar. Lett.* **31**, 51–64 (2011).
46. Dorale, J. A. et al. Sea-level highstand 81,000 years ago in Mallorca. *Science* **327**, 860–863 (2010).
47. Creveling, J. R., Mitrovica, J. X., Clark, P. U., Waelbroeck, C. & Pico, T. Predicted bounds on peak global mean sea level during marine isotope stages 5a and 5c. *Quat. Sci. Rev.* **163**, 193–208 (2017).
48. Pomar, L. Reef geometries, erosion surfaces and high-frequency sea-level changes, upper Miocene Reef Complex, Mallorca, Spain. *Sedimentology* **38**, 243–269 (1991).
49. Pomar, L. La evolución tectonosedimentaria de las Baleares: análisis crítico. *Acta Geol. Hisp.* **14**, 293–310 (1979).
50. Miller, K. G. et al. The Phanerozoic record of global sea-level change. *Science* **310**, 1293–1298 (2005).
51. Flament, N. et al. A review of observations and models of dynamic topography. *Lithosphere* **5**, 189–210 (2013).
52. Stocchi, P. et al. MIS 5e relative sea-level changes in the Mediterranean Sea: contribution of isostatic disequilibrium. *Quat. Sci. Rev.* **185**, 122–134 (2018).
53. Meehl, G. A. et al. *Climate Change 2007: The Physical Science Basis* (Cambridge Univ. Press, 2007).

**Author contributions** B.P.O., J.J.F. and O.A.D. designed the research. A.G. and J.G. first documented the presence of POS in the cave studied. B.P.O., J.J.F., J.G., A.G. and O.A.D. measured POS elevations and collected samples. V.J.P., O.A.D. and Y.A. measured U–Pb ages of POS samples. J.A. produced the GIA and long-term deformation corrections. O.A.D., J.A. and B.P.O. drafted and wrote the majority of the manuscript with input from all authors.
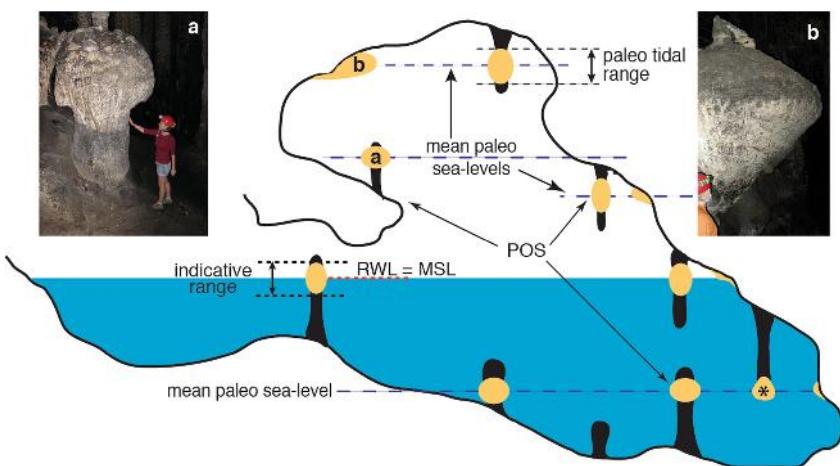
**Extended Data Fig. 1 | Schematic profile of a coastal cave in Mallorca hosting POS at different levels. a, b,** Standalone (**a**) and cave-wall (**b**) POS structures. The asterisked POS is an example of an asymmetric knob-like carbonate encrustation that forms when only the tip of the stalactite is submerged.
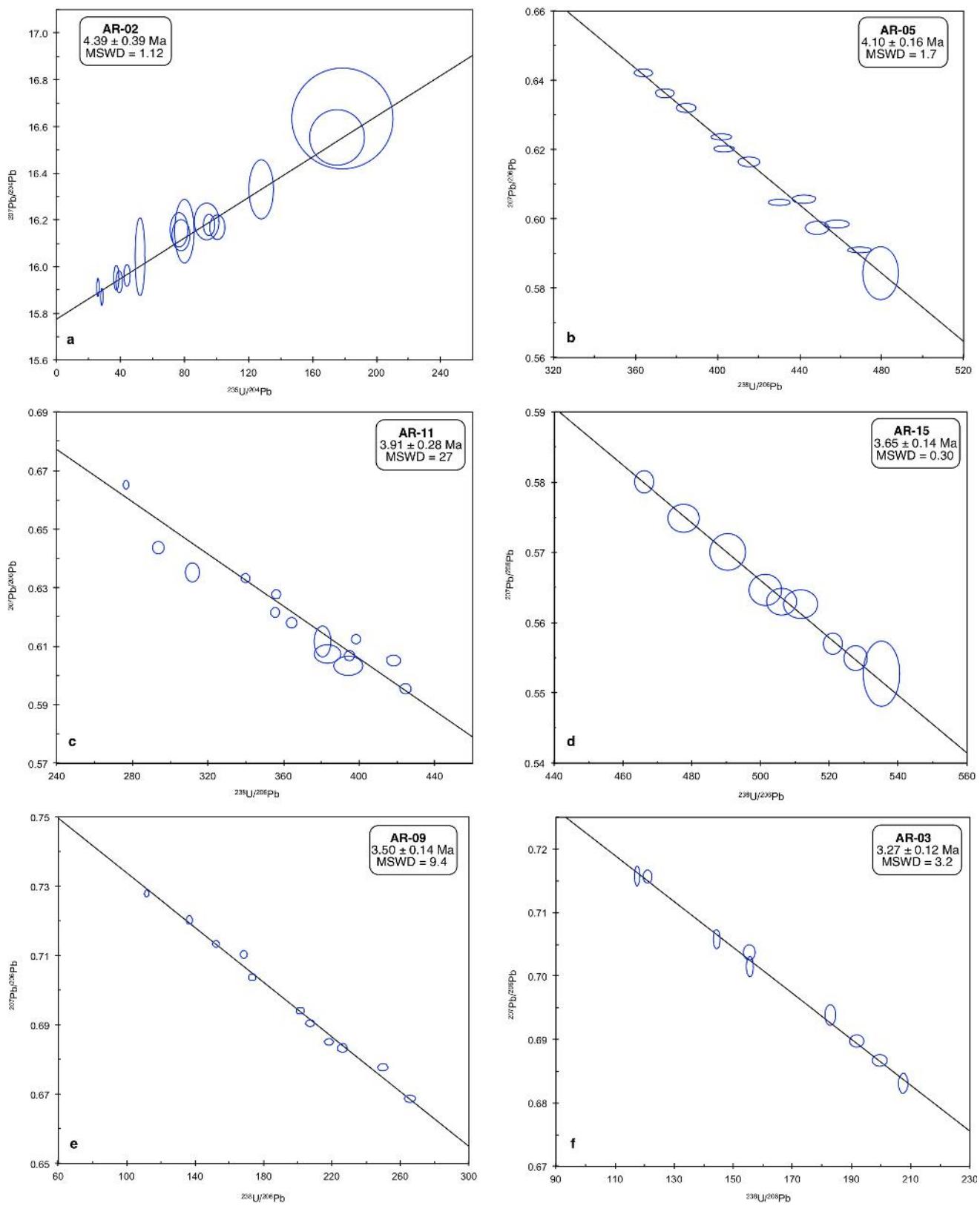
**Extended Data Fig. 2 | U–Pb isochrons and calculated age estimates for the six POS samples. a,** $^{235}U$–$^{207}Pb$ two-dimensional isochron for sample AR-02; **b–f**, Concordia-constrained linear three-dimensional isochron for samples AR-05 (**b**), AR-11 (**c**), AR-15 (**d**), AR-09 (**e**) and AR-03 (**f**). Error ellipses on individual ages are $2\sigma$.
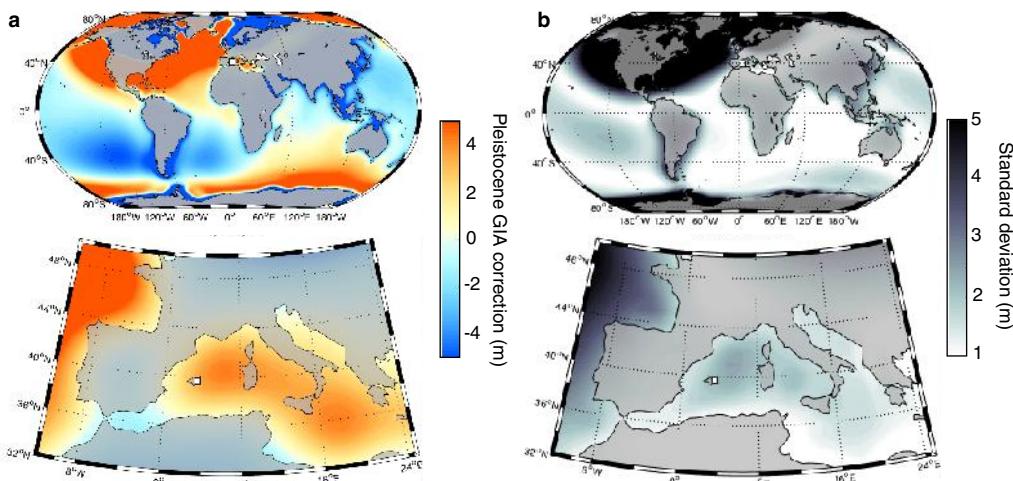
**Extended Data Fig. 3 | GIA contribution due to ongoing adjustment.**
This GIA contribution is caused by the incomplete present-day adjustment to the late Pleistocene ice and ocean loading cycles. **a**, Model simulation using a viscosity structure of $5 \times 10^{20}$ Pa s viscosity in the upper mantle, $5 \times 10^{21}$ Pa s viscosity in the lower mantle, and an elastic lithospheric thickness of 96 km. **b**, Standard deviation of model predictions obtained using 36 different radial viscosity profiles, including varying the lithospheric thickness. The square marks the position of Coves d'Artà. The figures were produced using Matlab 2015b and the m_map plotting package (https://www.eoas.ubc.ca/~rich/map.html).
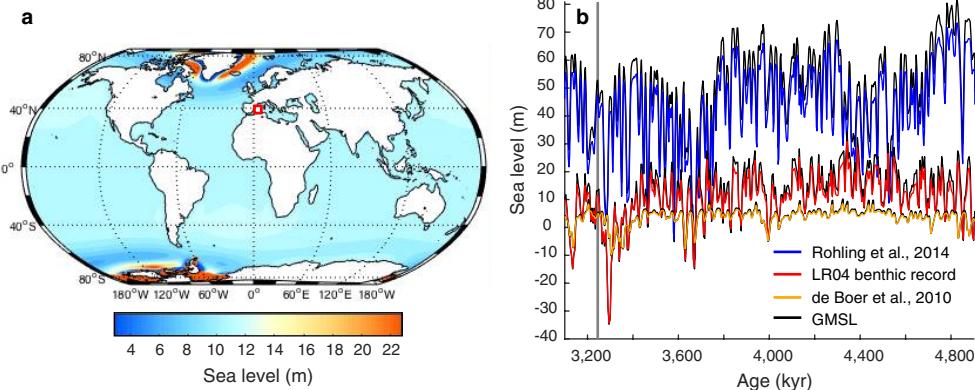
**Extended Data Fig. 4 | GIA contribution due to Pliocene ice age cycles.**
The model simulation uses a viscosity structure of $5 \times 10^{20}$ Pa s viscosity
in the upper mantle, $5 \times 10^{21}$ Pa s viscosity in the lower mantle, and an
elastic lithospheric thickness of 96 km. **a**, Snapshot of sea level at 3.244
Ma (grey vertical line in **b**) assuming a GMSL curve based on the LR04
benthic record[10]. The colour scale is chosen to diverge around the GMSL
value of 13 m. The red square marks the position of Coves d'Artà. The
figure was produced using Matlab 2015b and the m_map plotting package

(https://www.eoas.ubc.ca/~rich/map.html). **b**, Local sea level at Coves
d'Artà based on Rohling et al.[2] (blue), de Boer et al.[4] (yellow), and the
LR04 benthic record[10] (red). Respective GMSL curves are shown in black
and mostly coincide with local sea level at Coves d'Artà (note that for the
estimates based on Rohling et al.[2] the black GMSL curve is mostly behind
the local sea-level curve in blue). Sea level is relative to the beginning of
this run (4.9 Ma).
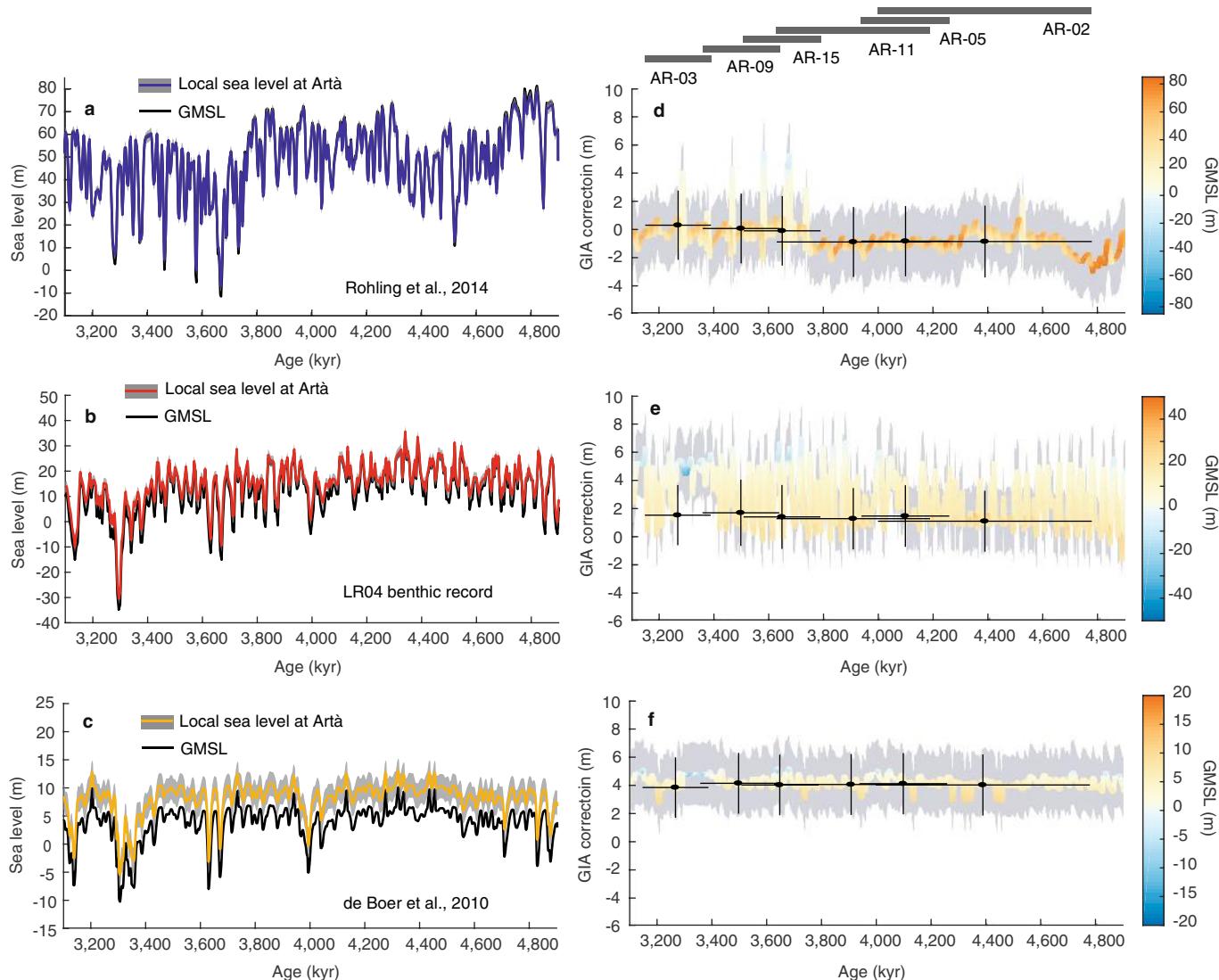
**Extended Data Fig. 5 | Total GIA correction. a–c,** Local sea-level change at Coves d'Artà, as calculated from the GIA model based on the GMSL curves by Rohling et al.[2] (**a**), the LR04 benthic record[10] (**b**) and de Boer et al.[4] (**c**). Uncertainties due to Earth's viscoelastic structure are denoted by grey bands. **d–f,** GIA correction colour coded by the GMSL value; standard deviations are shown as grey bands. Black markers indicate the GIA correction and its uncertainty for each POS. Results are for the GMSL curves by Rohling et al.[2] (**d**), the LR04 benthic record[10] (**e**) and de Boer et al.[4] (**f**).
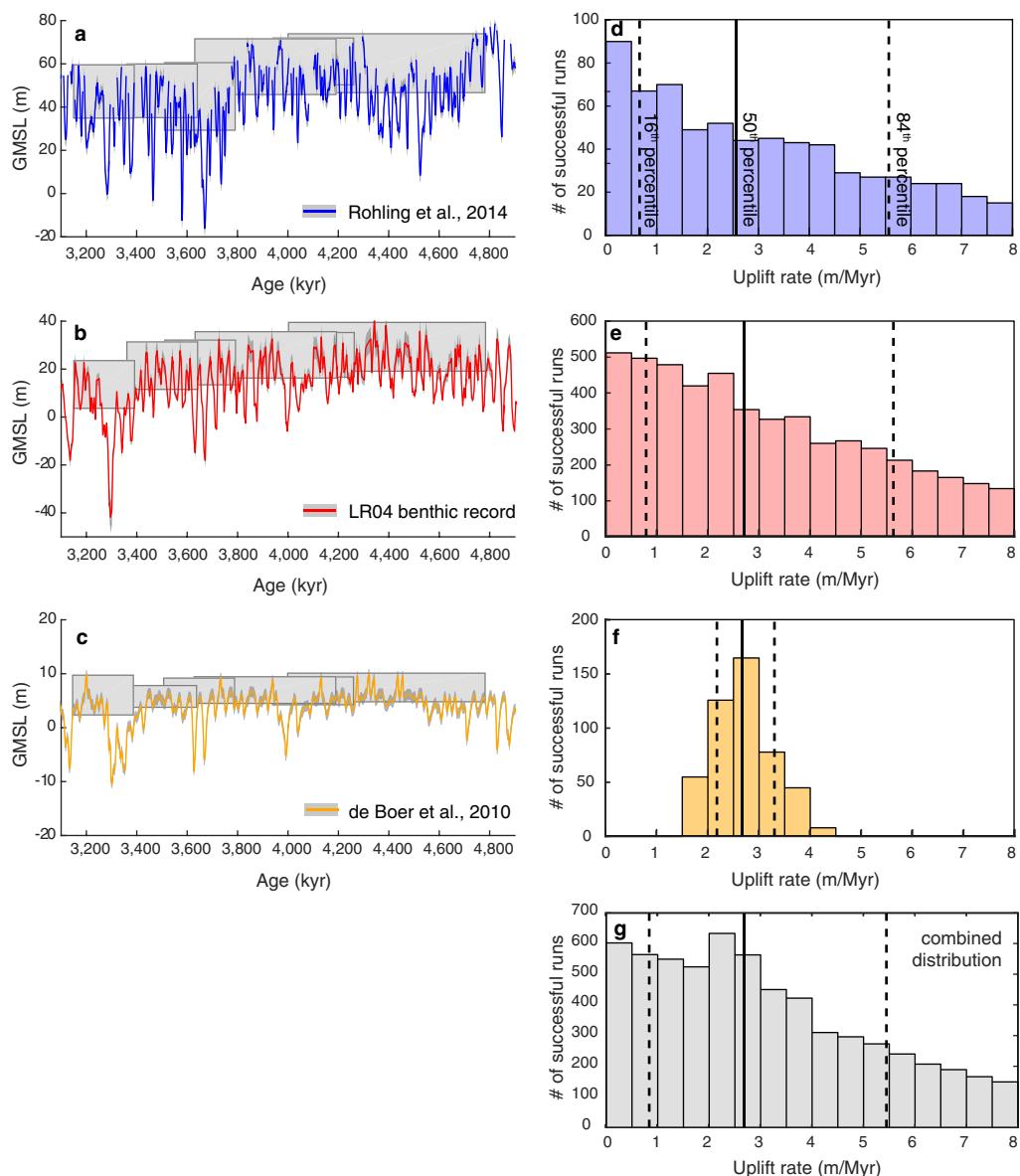
**Extended Data Fig. 6 | Uplift rate estimation.** Determining the amount of uplift based on the best fit of observed relative sea-level changes across the POS to other GMSL reconstructions over the same time interval. **a–c**, GMSL curves[2,4,10]; grey bars are $1\sigma$ uncertainties. Boxes indicate the age uncertainty for each POS and the 50th and 99th percentiles of the GMSL values that fall within this age range. We calculate synthetic sea-level changes relative to the youngest POS and compare them to the observed sea-level changes, assuming a range of uplift rates. **d–f**, Histograms of uplift rates in which we find a good fit between the observed and the synthetic data. Percentiles (16th, 50th and 84th) are shown by vertical lines (solid line is the median, dashed lines are uncertainty bounds). We conducted ten million iterations for this Monte Carlo search. **g**, Histogram combining all uplift rates that resulted in a good fit.
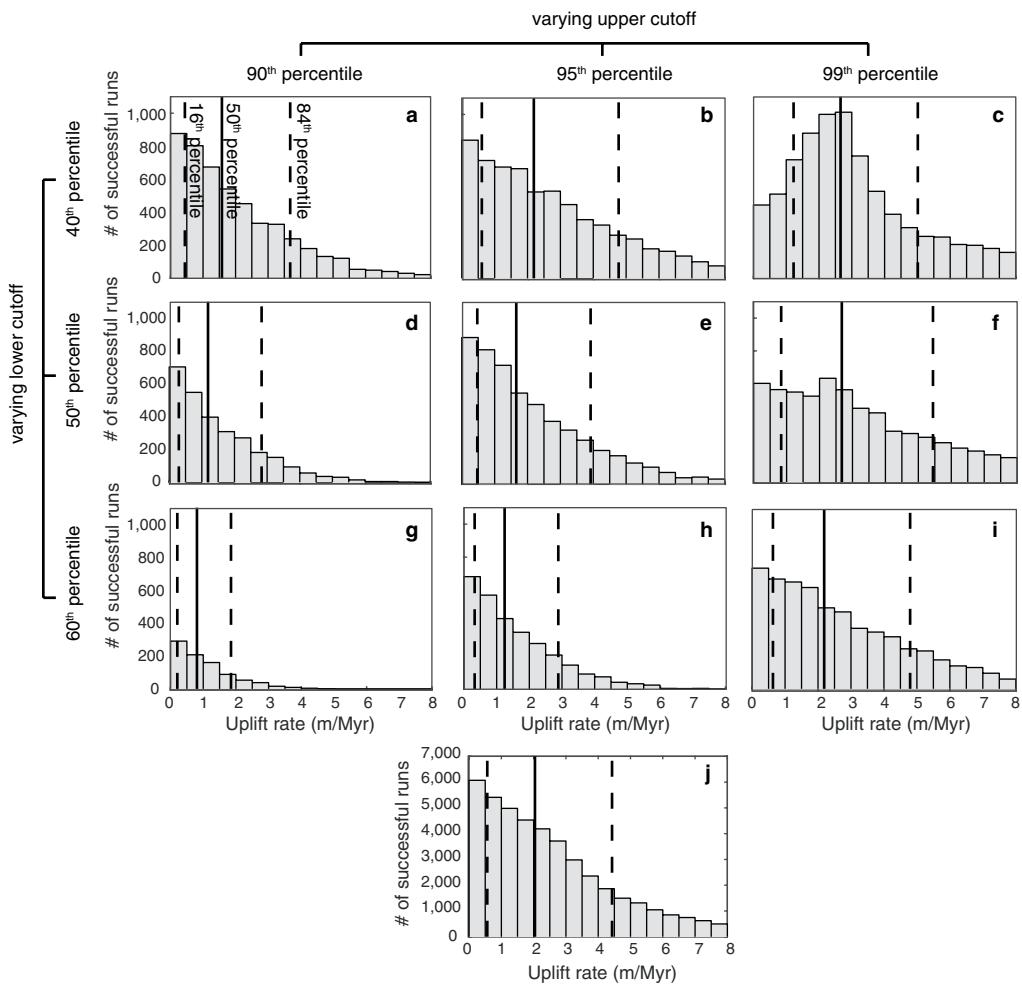
**Extended Data Fig. 7 | Joint uplift rate estimation from varying the upper and lower cutoff. a–i,** Joint histograms for a variety of lower and upper percentile cutoffs. The lower cutoff was varied between the 40th, 50th and 60th percentiles (different rows), whereas the upper cutoff was varied between the 90th, 95th and 99th percentiles (different columns). Percentiles in the histograms (16th, 50th and 84th) are shown by vertical lines (solid line is the median, dashed lines are uncertainty bounds). Panel **f** is identical to Extended Data Fig. 6g. **j,** Combination of all joint histograms to obtain our best-estimate uplift rate used for Table 1, Fig. 2c and Fig. 3.
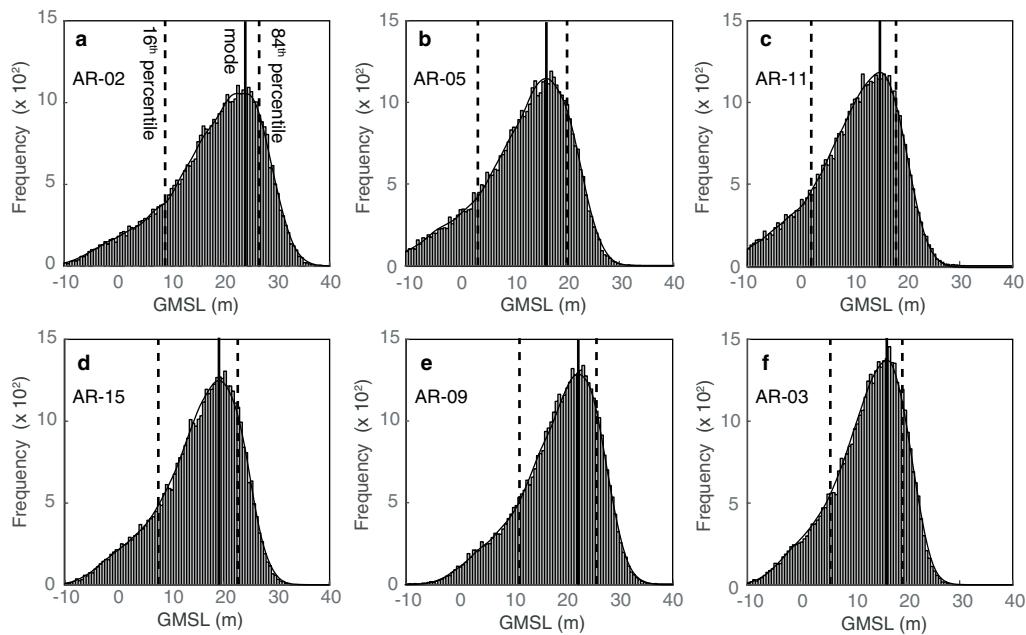
**Extended Data Fig. 8 | Reconstructed elevation of the GMSL for each POS after all corrections have been applied. a–f,** PDFs for the GMSL estimate for AR-02, AR-05, AR-11, AR-15, AR-09 and AR-03, respectively. PDFs are constructed assuming Gaussian uncertainties for the measured elevation of POS, the respective indicative range, the GIA correction, thermal expansion and POS age (Table 1); a non-Gaussian distribution is obtained for the uplift correction (Table 1, Extended Data Figs. 6, 7). The mode (thick black line) and lower and upper uncertainty bounds (16th and 84th percentiles, thick dashed lines) are shown by vertical lines and correspond to the GMSL estimates reported in Table 1. We used a kernel with 1-m bandwidth to calculate the mode and PDF (thin black line).

**Extended Data Table 1 | GMSL estimates for different percentiles (10th, 33th, 50th, 66th and 90th) following IPCC's likelihood scale**

| Sample code | Age (Myr) | GMSL (m) | | | | | GMSL without correction for thermal expansion (m) | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | 10th | 33th | 50th | 66th | 90th | 10th | 33th | 50th | 66th | 90th |
| AR-02 | 4.39 ± 0.39 | 5.3 | 15.5 | 19.8 | 23.1 | 28.2 | 6.9 | 17.1 | 21.4 | 24.7 | 29.8 |
| AR-05 | 4.10 ± 0.16 | 0.1 | 9.5 | 13.6 | 16.7 | 21.7 | 1.6 | 11.0 | 15.1 | 18.2 | 23.2 |
| AR-11 | 3.91 ± 0.28 | -1.3 | 7.8 | 11.6 | 14.6 | 19.5 | 0.1 | 9.2 | 13.0 | 16.0 | 20.9 |
| AR-15 | 3.65 ± 0.14 | 4.5 | 13.1 | 16.6 | 19.4 | 24.0 | 5.8 | 14.4 | 17.9 | 20.7 | 25.3 |
| AR-09 | 3.50 ± 0.14 | 8.2 | 16.5 | 19.9 | 22.6 | 27.0 | 9.4 | 17.7 | 21.1 | 23.8 | 28.2 |
| AR-03 | 3.27 ± 0.12 | 2.9 | 10.5 | 13.7 | 16.3 | 20.4 | 4.1 | 11.7 | 14.9 | 17.5 | 21.6 |

# LETTER

# The amplitude and origin of sea-level variability during the Pliocene epoch

G. R. Grant[1,2]*, T. R. Naish[1], G. B. Dunbar[1], P. Stocchi[3], M. A. Kominz[4], P. J. J. Kamp[5], C. A. Tapia[6], R. M. McKay[1], R. H. Levy[1,2] & M. O. Patterson[7]

Earth is heading towards a climate that last existed more than three million years ago (Ma) during the 'mid-Pliocene warm period'[1], when atmospheric carbon dioxide concentrations were about 400 parts per million, global sea level oscillated in response to orbital forcing[2,3] and peak global-mean sea level (GMSL) may have reached about 20 metres above the present-day value[4,5]. For sea-level rise of this magnitude, extensive retreat or collapse of the Greenland, West Antarctic and marine-based sectors of the East Antarctic ice sheets is required. Yet the relative amplitude of sea-level variations within glacial–interglacial cycles remains poorly constrained. To address this, we calibrate a theoretical relationship between modern sediment transport by waves and water depth, and then apply the technique to grain size in a continuous 800-metre-thick Pliocene sequence of shallow-marine sediments from Whanganui Basin, New Zealand. Water-depth variations obtained in this way, after corrections for tectonic subsidence, yield cyclic relative sea-level (RSL) variations. Here we show that sea level varied on average by 13 ± 5 metres over glacial–interglacial cycles during the middle-to-late Pliocene (about 3.3–2.5 Ma). The resulting record is independent of the global ice volume proxy[3] (as derived from the deep-ocean oxygen isotope record) and sea-level cycles are in phase with 20-thousand-year (kyr) periodic changes in insolation over Antarctica, paced by eccentricity-modulated orbital precession[6] between 3.3 and 2.7 Ma. Thereafter, sea-level fluctuations are paced

by the 41-kyr period of cycles in Earth's axial tilt as ice sheets stabilize on Antarctica and intensify in the Northern Hemisphere[3,6]. Strictly, we provide the amplitude of RSL change, rather than absolute GMSL change. However, simulations of RSL change based on glacio-isostatic adjustment show that our record approximates eustatic sea level, defined here as GMSL unregistered to the centre of the Earth. Nonetheless, under conservative assumptions, our estimates limit maximum Pliocene sea-level rise to less than 25 metres and provide new constraints on polar ice-volume variability under the climate conditions predicted for this century.

Highly resolved climate and sea-level reconstructions from the Pliocene provide insights into the response of the polar ice sheets to climate forcings projected for the twenty-first century[1,5]. For example, while it is acknowledged that palaeogeography was subtly different from that of the present day, polar ice-sheet configuration geometry was broadly similar, and therefore computer simulations of ice sheets can be used to constrain the equilibrium response of global sea level to $CO_2$ partial pressures of[5,7] 350–400 ppm. Pliocene sea-level changes have been reconstructed using a variety of geological techniques including: (i) marine benthic oxygen-isotope ($\delta^{18}O$) records paired with Mg/Ca palaeothermometry (a proxy for global ice volume)[4], (ii) an algorithm incorporating sill-depth, salinity and the $\delta^{18}O$ record from the Mediterranean and Red seas[8], (iii) uplifted palaeo-shorelines[4,9], and (iv) backstripped continental margins[2,4]. In addition to the considerable
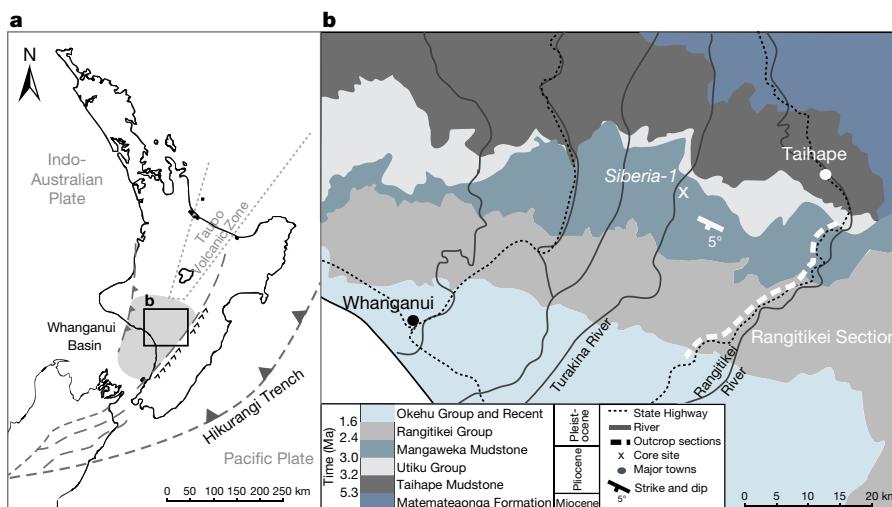


**Fig. 1 | Location of Whanganui Basin, New Zealand, and sample sites.**
**a**, Overview of North Island. Whanganui Basin (grey shaded region) formed behind the Hikurangi subduction zone as part of a southward-migrating pattern of lithospheric flexure associated with southwestward propagation of the subducting Pacific Plate beneath the Indo-Australian Plate[2]. **b**, Magnified view of boxed area in **a**. Subsequent uplift in central

North Island during the last 1 Ma, in response to redistribution of lithosphere over the mantle[2], has exposed Plio-Pleistocene, shallow-marine sediments onshore where they tilt southwestward at 5°. Locations of Siberia-1 drill site (white 'x' marker) and Rangitikei River outcrop (bold dashed white line) are shown. Geological data in **b** adapted from GNS Science.

[1]Antarctic Research Centre, Victoria University of Wellington, Wellington, New Zealand. [2]GNS Science, Lower Hutt, New Zealand. [3]Coastal Systems Department, Royal Netherlands Institute for Sea Research, and Utrecht University, Den Burg, The Netherlands. [4]Department of Geological and Environmental Sciences, Western Michigan University, Kalamazoo, MI, USA. [5]School of Science, University of Waikato, Hamilton, New Zealand. [6]Departamento de Obras Civiles y Geologia, Facultad de Ingenieria, Universidad Catolica de Temuco, Temuco, Chile. [7]Binghamton University, State University of New York, Binghamton, NY, USA. *e-mail: g.grant@gns.cri.nz
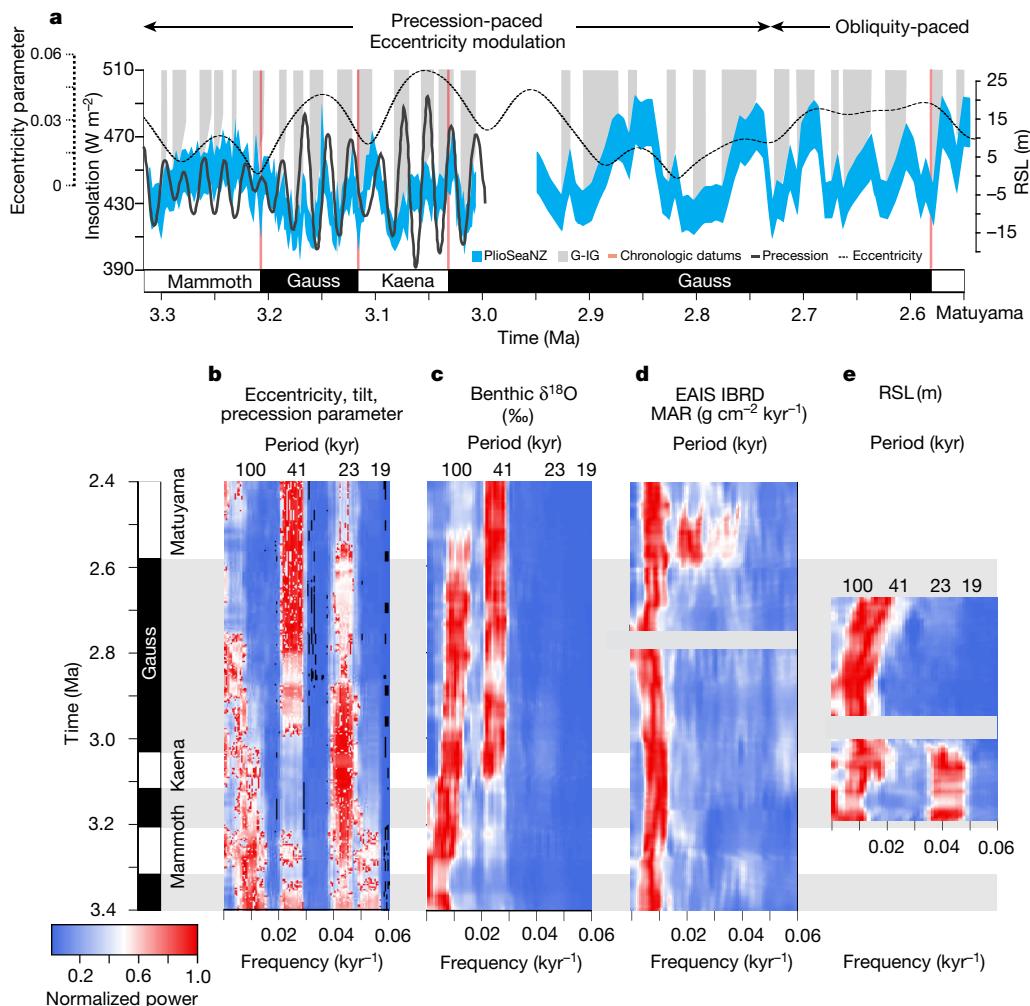
**Fig. 2 | The PlioSeaNZ RSL record and comparison with orbital parameters and climate proxies. a**, PlioSeaNZ RSL record (right-hand vertical axis), unregistered to the present day, for the middle to late Pliocene, with uncertainty represented by the shaded blue band, which does not exceed ±5.6 m (see Methods). Glacial–interglacial (G–IG) transitions are marked by the shaded grey bands. The age model is untuned and derived from linear sedimentation rates between magnetic reversals (orange-pink lines) with an uncertainty of ±5 kyr. Summer insolation (1 January) at 65° S (black curve) and the eccentricity parameters (dashed curve) are shown for ref. [18] (left-hand vertical axes). **b–e**, Multi-taper method time–frequency analyses (see Methods) displaying normalized power (colour scale) for eccentricity, obliquity and precession insolation parameter[18] (**b**), the global benthic foraminifera $\delta^{18}O$ stack[3] (**c**), EAIS IBRD mass accumulation rate[21] (**d**) and our RSL record (**e**; PlioSeaNZ Whanganui Basin). Periods are denoted for eccentricity (100 kyr), obliquity (41 kyr) and precession (23 and 19 kyr).

uncertainties associated with these techniques[1], reconstructing peak Pliocene GMSL, with respect to the present day, is hindered by the influence of Earth deformation processes, which can cause local sea-level changes as large as the ice-volume contribution. Global mantle dynamic processes have caused vertical land movement of tens of metres since the Pliocene[9]. The viscoelastic response of the crust and rotational and gravitational changes, known collectively as glacio-isostatic adjustment (GIA), occur due to redistribution of water between ice sheets and the oceans, and can cause substantial deviations from GMSL for sites in the near fields of ice sheets[10]. These processes cast considerable doubt on our ability to estimate peak Pliocene GMSL, registered to the present day, using established methodologies[1,5,9,10].

Although the global benthic $\delta^{18}O$ stack provides one of the most detailed proxies for orbital-scale (glacial–interglacial) climate variability during the Pliocene[3], the signal comprises both ocean-temperature and ice-volume effects that are not easily deconvolved[2,4,11]. Moreover, calibrations of $\delta^{18}O$ to sea level do not account for the nonlinear relationship between marine-based ice-volume change and the $\delta^{18}O$ of sea water[12].

Shallow-marine continental margin sediments contain a range of biological and sedimentological tracers sensitive to multi-metre water depth changes, and thus far offer the greatest potential for accurate reconstruction of Pliocene sea-level change on orbital timescales[2,4]. However, their interpretation has been hampered by the limited precision inherent in foraminiferal palaeo-depth indicators and the influence of sediment erosion during sea-level lowstands, which preclude the resolution of the full amplitude of sea-level change. Whanganui Basin, New Zealand (Fig. 1), provides a sedimentary fill about 5 km thick, accumulated under relatively linear basin subsidence due to plate boundary interactions behind the Hikurangi subduction zone, off eastern New Zealand (Fig. 1), and offers one of the highest-resolution shallow-marine records of orbitally paced, Late Neogene global sea-level change in the world[2].

Here we reconstruct the amplitude and frequency of global glacial–interglacial scale sea-level changes between 3.3 and 2.5 Ma. We account for GIA and discount the effect of dynamic topography in assessing the relative amplitude on glacial–interglacial timescales as we do not attempt to register sea-level variations to the present day.

Our record, which we term PlioSeaNZ, is constructed from sedimentary cycles that represent fluctuations between middle- to outer-shelf water depths that were recovered in sediment cores (3.3–3.0 Ma) and outcrop sections exposed in the Rangitikei River valley (2.9–2.5 Ma). Sediments accumulated continuously at rates of >1 m kyr$^{-1}$ (see Methods). Erosion during lowstands did not occur on the middle
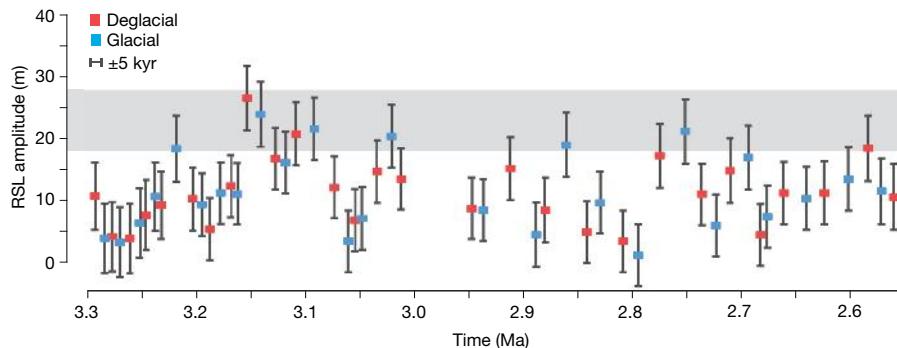
**Fig. 3 | Amplitude of orbital-scale, sea-level fluctuations in the PlioSeaNZ RSL record.** Amplitudes of deglacial (glacial–interglacial; pink squares; $n = 28$) and glacial (interglacial–glacial; blue squares; $n = 26$) RSL changes are shown with error bars representing $\pm 1$ s.d. (after equation (10)) with an average of 5.1 m, and age uncertainty is $\pm 5$ kyr (as discussed in the text, and shown in the figure key). The grey shaded band (about $23 \pm 5$ m) shows the possible contribution from the marine-based sectors of the AIS (about 23 m)[28] and the GIS estimated as[19] $\pm 5$ m depending on the interhemispheric phase relationship. Glacial–interglacial amplitudes higher than approximately 28 m exceed the ice inventory of the marine-based AIS sectors (22.7 m; ref. [28]) and the GIS (5 m; ref. [19]) based on present-day volumes.

to outer shelf, because the changes in the amplitude of Pliocene RSL were accommodated in these environments without experiencing wave base erosion or subaerial exposure. The palaeo-environmental interpretation of the cores and outcrops is described in detail in ref. [6] and summarized in Supplementary Figs. 1 and 2. Near-synchronous vertical changes in sediment grain size, lithofacies and benthic foraminiferal assemblages support the interpretation that sediment was deposited in hydrodynamic equilibrium with the contemporary wave climate[6]. Furthermore, an in-phase relationship between climate and water-depth variability was established on the basis of coeval changes in fossil pollen assemblages and sediment grain size[6]. The age model used here is based on a new high-resolution magnetostratigraphy for the core[13] and existing magnetostratigraphy for the outcrop, calibrated with biostratigraphy and tephrochronology[6].

We have developed a novel approach that utilizes the well-established relationship between sediment grain size and water depth[14] to calculate palaeo-water-depth changes. Wave energy produces a decreasing near-bed velocity at increasing water depths across the shelf, resulting in a seaward-fining sediment profile[14]. Modern observations support theoretical calculations that show that maximum water depth for a given grain size corresponds to the depth at which wave-induced near-bed velocity exceeds the critical velocity required for sediment transport[14] (see Methods; Extended Data Fig. 1a). Thus, the percentage of sand (grains of size 63–2,000 μm) in closely spaced geological samples can be used to estimate changes in palaeo-water depth provided that the sediment is wave-graded[6] and that Pliocene wave climate can be broadly estimated.

Pliocene palaeogeographic reconstructions indicate that the Whanganui Basin was a west-facing embayment (Extended Data Fig. 2)[15], much like today. A similar modern wave climate is presumed for the Pliocene, as the primary influence on wave generation is fetch (the distance and time over which wind interacts with the sea surface), which has not changed[15]. Although global climate model simulations of the Pliocene show an approximately 2° poleward shift of the zonal westerlies compared with the Holocene, wind speed variance is[16] only $\pm 0.5$ m s$^{-1}$. Sensitivity analysis for a range of wave heights in response to wind speed variance reveals a maximum 0.5-m variation between resulting water depth versus grain size profiles (see Methods; Extended Data Fig. 1b). Therefore, the analysis presented here is based on the modern wave climate, which is most plausible on the basis of this limited information.

A two-dimensional backstripping method[16] is applied to the resulting local water depths to remove the effects of tectonic subsidence and compaction caused by sediment and water loading (see Methods). The resulting RSL curve, the PlioSeaNZ record, provides new constraints on the amplitude and frequency of glacial–interglacial sea-level change between about 3.3 and 2.5 Ma (Fig. 2). The amplitude of deglacial and glacial phases of the PlioSeaNZ curve ($n = 54$) are shown in Fig. 3. Uncertainty estimates for individual amplitude changes vary, and are outlined in Methods and Extended Data Table 1, but do not exceed $\pm 5.6$ m. The average amplitude is $13 \pm 5$ m ($n = 44$), excluding phases for which uncertainty exceeds amplitude ($<5$ m; $n = 10$). Between 3.3 and 2.7 Ma, the duration of glacial–interglacial cyclicity is about 20 kyr within a longer 100-kyr envelope, which spans a stratigraphic break (approximately 3.0–2.95 Ma), that represents correlation to the Rangitikei section, and is consistent with pacing by eccentricity-modulated orbital precession (Fig. 2b)[6]. Thereafter, sea-level fluctuations are paced by cycles of 41 kyr, corresponding to changes in Earth's axial tilt, with a diminishing influence of precession (Fig. 2e).

Three magnetic reversals are correlated to the top of the Mammoth Subchron and the base and top of the Kaena Subchron (3.3 to 3.0 Ma)[6]. The age model derived from linear interpolation between magneto-stratigraphic reversals (sampled every 1 kyr approximately), is not orbitally tuned. These age datums offer sufficient precision (about $\pm 5$ kyr) to resolve the phasing of sea-level variability with respect to the timing of astronomical forcing[17,18] (Fig. 2a). Results show that precession-paced sea-level cycles are positively correlated with southern high-latitude summer (65° S, 1 January) insolation and negatively correlated with Northern Hemisphere summer (65° N, 1 July) insolation (Extended Data Fig. 3). The phase coherence between the PlioSeaNZ record and Southern Hemisphere insolation implies a dominant Antarctic Ice Sheet (AIS) meltwater source. We modelled the glacio-isostatic adjustment (GIA) associated with a plausible range of Pliocene glacial–interglacial polar ice volumes[7,19,20] and Earth viscosity models (see Methods). The resulting ensemble of geoidal sea-level changes shows that the PlioSeaNZ record is within $\pm 5\%$ of eustatic sea level (ESL), which equates to $\pm 0.75$–1.25 m for the three ESL scenarios examined (Fig. 4; see Methods). Whereas error in RSL estimates derives directly from the water depth–grain size model (Fig. 3), conversion to ESL adds $\pm 1.25$–1.75 m to the uncertainty owing to the combined influence of wave climate variability and GIA deviation, which are dependent on the respective wind-field and meltwater scenarios used.

Our results for the interval from 3.2 to 2.7 Ma are consistent with precession dominance in an iceberg-rafted debris (IBRD) sedimentary record of East Antarctic Ice Sheet (EAIS) dynamics[21], and weak obliquity in Arctic IBRD records[22] and the global benthic $\delta^{18}$O stack[3] (Fig. 2). Although the Arctic IBRD record[22] indicates that marine calving ice sheets occupied Greenland at this time and some ice-sheet simulations[19] suggest that precession-paced anti-phase bi-polar ice-volume change occurred through this interval, our results preclude an anti-phase Northern Hemisphere melt water contribution beyond that of the Greenland Ice Sheet (GIS), especially between[23] 3.3 and 3.0 Ma. The dominance of 41-kyr variability in the PlioSeaNZ record after about 2.7 Ma is consistent with increased obliquity signals recorded
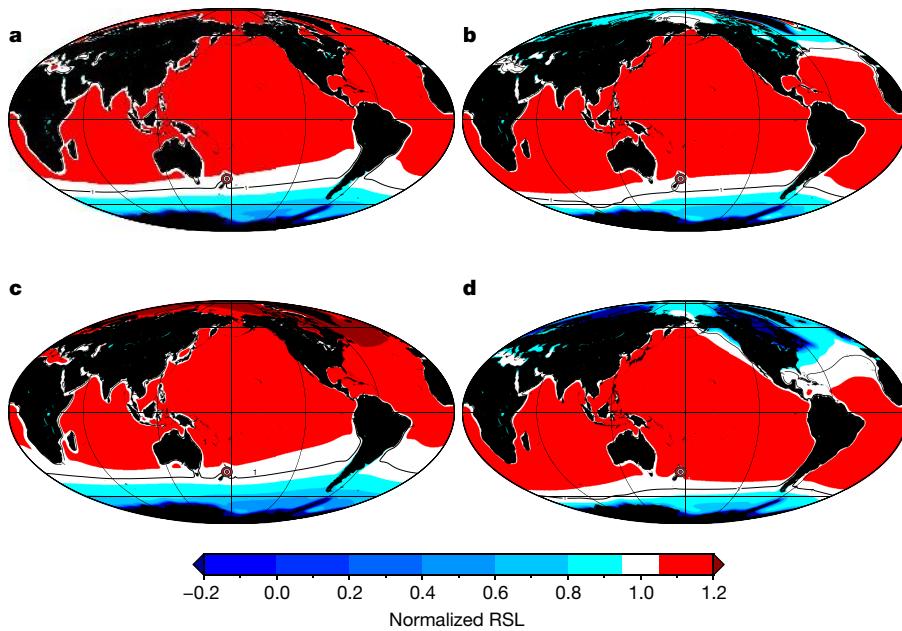
**Fig. 4 | The pattern of normalized RSL from GIA simulations of a 20-m rise in ESL.** Modelled result of 10 kyr linear melting between glacial and interglacial states required for 20 m of equivalent ESL, and according to the reference mantle viscosity profile[30]. **a**, 20 m ESL released from AIS only. **b**, AIS and GIS synchronously release 15 m and 5 m ESL, respectively. **c**, AIS releases 25 m ESL while GIS accumulates 5 m ESL (that is, in anti-phase). **d**, AIS and NHIS synchronously release 10 m ESL. The white band represents ±0.05 of the eustatic mean (bold black line), which equates to ±1 m. The Whanganui site is highlighted by the red and white bullseye on New Zealand.

in global ice volume[3] (Fig. 2), and North Atlantic[24] and tropical[25] sea surface temperature records. This timing also coincides with changes in Southern Hemisphere zonal wind strength[26], global cooling and the intensification of Northern Hemisphere glaciation[3]. Intriguingly, the amplitude of late Pliocene 41-kyr-duration sea-level changes in the PlioSeaNZ record remains comparatively low, suggesting that major continental-scale glaciation in the Northern Hemisphere did not occur until approximately 2.5 Ma, as originally proposed in ref. [27]. These low amplitudes could be a result of obliquity-paced intensification of Northern Hemisphere ice sheets (NHIS) masked by an anti-phase response (every second precession cycle) of the AIS.

Today, regions of the AIS grounded below sea level contain an ice volume of about 22.7 m sea-level equivalent (SLE)[28]. We suggest that Pliocene sea-level variations with glacial–interglacial amplitudes of less than about 23 m only require fluctuations of Antarctica's marine-based ice, and do not require contribution from NHIS. An (incomplete) deglaciation of Greenland of 5 m SLE, in phase with marine-based ice-sheet retreat in Antarctica, can explain RSL variations up to +28 m. But if these deglaciations were anti-phased, the maximum amplitude of glacial–interglacial sea-level change above the present-day value would be 18 m (Fig. 3). A larger than present-day AIS during Pliocene glacials, which is supported by geological evidence[29], could explain amplitudes exceeding 28 m.

The average magnitude of glacial–interglacial sea-level variability in the PlioSeaNZ record (13 ± 5 m) between 3.3 and 3.0 Ma reflects melt-water originating from a highly sensitive AIS that regularly retreated and advanced in response to changes in southern high-latitude insolation before intensification of the NHIS. The proximal Antarctic IBRD record from IODP Site U1361 adjacent to the Wilkes subglacial basin[21] suggests that marine-based portions of the EAIS continued to respond to precession until about 2.5 Ma (Fig. 2), albeit with decreasing amplitude. Therefore, we attribute the 41-kyr obliquity forcing in our record after about 2.7 Ma to the intensification of NHIS. This, together with reduced amplitude variance in the AIS, both in-phase and anti-phase with NHIS, was driving global sea-level fluctuations up to an amplitude of approximately 25 m (Fig. 3).

In conclusion, our results provide new constraints on polar ice-sheet and global sea-level variability during the middle and late Pliocene, that

are: (i) independent of estimates from the global benthic $\delta^{18}O$ stack[3] and other geochemical proxies[4], and (ii) broadly consistent with AIS models[7,19,20] that simulate a contribution of 13–17 m to global sea-level rise above present. Because our record cannot be registered to present-day sea level, we cannot directly constrain the magnitude of peak Pliocene GMSL above present. Regardless, our results provide key insights into AIS sensitivity when Earth's climate equilibrates at a $CO_2$ partial pressure of about 400 ppm. Furthermore, if all the variability in the PlioSeaNZ record was above present-day sea level, then GMSL during the warmest mid-Pliocene interglacial was no more than +25 m. Although ice-sheet, ocean and continental geometries were subtly different during the mid-Pliocene, our results suggest that major loss of Antarctica's marine-based ice sheets, and an associated GMSL rise of up to 23 m, is likely if $CO_2$ partial pressures remain above 400 ppm.

## Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-019-1619-z.

1. Masson-Delmotte, V. et al. in *Climate Change 2013: The Physical Science Basis. Contribution of Working Group 1 to the Fifth Assessment Report of Intergovernmental Panel on Climate Change* (eds Stocker, T. F. et al.) Ch. 5 (Cambridge Univ. Press, 2013).
2. Naish, T. R. & Wilson, G. S. Constraints on the amplitude of Mid-Pliocene (3.6–2.4 Ma) eustatic sea-level fluctuations from the New Zealand shallow-marine sediment record. *Phil. Trans. R. Soc. Lond. A* **367**, 169–187 (2009).
3. Lisiecki, L. E. & Raymo, M. E. A Pliocene-Pleistocene stack of 57 globally distributed benthic $\delta^{18}O$ records. *Paleoceanography* **20**, PA1003 https://doi.org/10.1029/2004PA001071 (2005).
4. Miller, K. G. et al. High tide of the warm Pliocene: implications of global sea level for Antarctic deglaciation. *Geology* **40**, 407–410 (2012).
5. Dutton, A. et al. Sea-level rise due to polar ice-sheet mass loss during past warm periods. *Science* **349**, aaa4019 (2015).
6. Grant, G. et al. Mid- to late Pliocene (3.3–2.6 Ma) global sea-level fluctuations recorded on a continental shelf transect, Whanganui Basin, New Zealand. *Quat. Sci. Rev.* **201**, 241–260 (2018).

7. Pollard, D., DeConto, R. M. & Alley, R. B. Potential Antarctic Ice Sheet retreat driven by hydrofracturing and ice cliff failure. *Earth Planet. Sci. Lett.* **412**, 112–121 (2015).
8. Rohling, E. J. et al. Sea-level and deep-sea-temperature variability over the past 5.3 million years. *Nature* **508**, 477–482 (2014); erratum **510**, 432 (2014).
9. Rovere, A. et al. The Mid-Pliocene sea-level conundrum: glacial isostasy, eustasy and dynamic topography. *Earth Planet. Sci. Lett.* **387**, 27–33 (2014).
10. Raymo, M. E., Mitrovica, J. X., O'Leary, M. J., DeConto, R. M. & Hearty, P. J. Departures from eustasy in Pliocene sea-level records. *Nat. Geosci.* **4**, 328–332 (2011).
11. Evans, D., Brierley, C., Raymo, M. E., Erez, J. & Müller, W. Planktic foraminifera shell chemistry response to seawater chemistry: Pliocene–Pleistocene seawater Mg/Ca, temperature and sea level change. *Earth Planet. Sci. Lett.* **438**, 139–148 (2016).
12. Gasson, E., DeConto, R. M. & Pollard, D. Modeling the oxygen isotope composition of the Antarctic ice sheet and its significance to Pliocene sea level. *Geology* **44**, 827–830 (2016).
13. Tapia, C. A. et al. High-resolution magnetostratigraphy of mid-Pliocene (3.3–3.0 Ma) shallow-marine sediments, Whanganui Basin, New Zealand. *Geophys. J. Int.* **217**, 41–57 (2019).
14. van Rijn, L. C. Unified view of sediment transport by currents and waves. I: Initiation of motion, bed roughness, and bed-load transport. *J. Hydraul. Eng.* **133**, 649–667 (2007).
15. Trewick, S. A. & Bland, K. J. Fire and slice: palaeogeography for biogeography at New Zealand's North Island/South Island juncture. *J. R. Soc. NZ* **42**, 153–183 (2012).
16. Kominz, M. & Pekar, S. Oligocene eustasy from two-dimensional sequence stratigraphic back-stripping. *Bull. Geol. Soc. Am.* **113**, 291–304 (2001).
17. Lourens, L. J. et al. Evaluation of the Plio-Pleistocene astronomical timescale. *Paleoceanogr. Paleoclim.* **11**, 391–413 (1996).
18. Laskar, J. et al. A long-term numerical solution for the insolation quantities of the Earth. *Astron. Astrophys.* **428**, 261–285 (2004).
19. de Boer, B., Haywood, A. M., Dolan, A. M., Hunter, S. J. & Prescott, C. L. The transient response of ice volume to orbital forcing during the warm Late Pliocene. *Geophys. Res. Lett.* **44**, 10486–10494 (2017).
20. Golledge, N. et al. Antarctic climate and ice sheet configuration during the early Pliocene interglacial 4.23 Ma. *Clim. Past* **13**, 959–975 (2017).
21. Patterson, M. et al. Orbital forcing of the East Antarctic Ice Sheet during the Pliocene and Early Pleistocene. *Nat. Geosci.* **7**, 841–847 (2014).
22. Jansen, E., Fronval, T., Rack, F. & Channell, J. E. Pliocene-Pleistocene ice rafting history and cyclicity in the Nordic Seas during the last 3.5 Myr. *Paleoceanography* **15**, 709–721 (2000).
23. Bailey, I. et al. An alternative suggestion for the Pliocene onset of major northern hemisphere glaciation based on the geochemical provenance of North Atlantic Ocean ice-rafted debris. *Quat. Sci. Rev.* **75**, 181–194 (2013).
24. Lawrence, K. T., Herbert, T. D., Brown, C. M., Raymo, M. E. & Haywood, A. M. High-amplitude variations in North Atlantic sea surface temperature during the early Pliocene warm period. *Paleoceanogr. Paleoclim.* **24**, https://doi.org/10.1029/2008PA001669 (2009).
25. Herbert, T. D., Peterson, L. C., Lawrence, K. T. & Liu, Z. Tropical ocean temperatures over the past 3.5 million years. *Science* **328**, 1530–1534 (2010).
26. Martínez-Garcia, A. et al. Southern Ocean dust-climate coupling over the past four million years. *Nature* **476**, 312–315 (2011).
27. Shackleton, N. J. et al. Oxygen isotope calibration of the onset of ice-rafting and history of glaciation in the North Atlantic region. *Nature* **307**, 620 (1984).
28. Fretwell, P. et al. Bedmap2: Improved ice bed, surface and thickness datasets for Antarctica. *Cryosphere* **7**, 375–393 (2013).
29. Naish, T. et al. Obliquity-paced Pliocene West Antarctic ice sheet oscillations. *Nature* **458**, 322 (2009).
30. Spada, G. et al. Modeling Earth's post-glacial rebound. *Eos* **85**, 62–64 (2004).

## METHODS

**Stratigraphy and chronology.** The core (Siberia-1; 3.3–3.0 Ma; 39.6964° S, 175.5241° E)[6] and outcrop (Rangitikei River Section; about 3.0–2.5 Ma)[31] provide the composite stratigraphy for the PlioSeaNZ RSL curve (Supplementary Fig. 1). Siberia-1, continuously cored from 40 to 350 m, recovered middle- to outer-shelf environments of the Utiku Group. Fourteen cycles (numbered 1 to 14) are identified on the basis of alternating bioturbated mudstone lithologies with variable sand content ranging from 10% to 60%. The stratigraphically-younger Mangaweka Mudstone exposed in the Rangitikei River Section comprises highly bioturbated clay-rich mudstone to mudstone cycles (15–24) characterized by sand content ranging from 0% to 40%. These repeating lithologies, interpreted as shallowing (coarsening upwards) and deepening (fining upwards) facies successions, are consistent with changes in water-depth-related benthic foraminiferal assemblages. Deepest and shallowest water depths also correspond with warmest and coldest terrestrial climates (respectively) identified by co-registered pollen assemblages (Supplementary Fig. 2)[6]. The mid-Pliocene Siberia-1 core was analysed for grain size every ~2 m (149 samples) at a resolution of ~2 kyr. The late Pliocene Rangitikei River Section was analysed for grain size every ~6 m (69 samples) at a resolution of ~6 kyr (ref. [31]). Grain size was measured on an Aqueous Liquid Module of the Beckman-Coulter LS 13 320 Laser Diffraction Particle Size Analyser, following the removal of organics and calcium carbonate (materials not deposited in hydraulic equilibrium).

The age model[6,13] was developed from calibration of the magnetic polarity zonation stratigraphy to the Geomagnetic Polarity Timescale[32] by tephrochronology, tephrostratigraphy and biostratigraphy and assumes linear sedimentation rates between magnetic reversals[6,13,33], and is therefore untuned and independent of astronomical timescales[18] and the benthic $\delta^{18}O$ stack[3] (Extended Data Table 1). The data gap ~3–2.95 Ma represents a small interval of stratigraphic underlap between the Sibera-1 drill core (~3.3–3 Ma) and the Rangitikei River Section (~2.95–2.5 Ma).

**Palaeobathymetry.** Wave-graded shelves produce a seaward-fining grain size profile in response to decreasing wave-driven current velocity at depth[34,35]. A previous approach used to reconstruct palaeo-water depth used an empirical relationship between percentage mud and water depth for a modern wave-graded coast[36]. This observed relationship is modelled here by relating wave-induced velocity and velocity required to initiate sediment transport. We utilize the well-established linear wave theory related to near-bed current velocity[37] and the critical velocity required for grain movement by wave motion[14]. The utility of this approach for predicting the grain size–water depth relationship of wave-graded coasts was tested for three modern transects with varying wave climates (Extended Data Fig. 1a).

We use $D_{90}$ (diameter at 90% of the grain size histogram, where 0% is equal to the smallest grain size) as a grain size parameter rather than $D_{50}$ (median grain size; compare ref. [14]) (equation (1), see below), as we assume all non-cohesive sand is capable of being moved at a given depth under storm wave conditions (peak wave periods). $D_{90}$ provides a more suitable representation of sand preserved on the shelf, which is a product of time-averaged extreme wave climate and displays a well constrained linear relationship with sand percentage (Extended Data Fig. 1c). The best fit between modelled and observed water depth for the most densely sampled transect was found when the exponential function (0.43) for peak wave period ($T_p$) was substituted for the original value (0.33) in equation (1), which is shown in Extended Data Fig. 1d. Our approach and the use of $D_{90}$ is more sensitive at greater water depths compared with the mud percentage–water depth relationship[36] as the polynomial function used to relate percentage mud and water depth in ref. [36] is unconstrained where mud exceeds ~85% (that is, <15% sand).

A broad spectrum of wave climate parameters was initially used in calculating the critical velocity required for sediment transport by wave motion ($U_{cr,w}$ in m s$^{-1}$; equation (1))[14] for a range of grain sizes ($D_{90}$ in units of m; 50–500 μm at 1 μm increments) and peak wave periods ($T_p$ in units of s; 5–20 s at 1 s increments). The wave-induced near-bed velocity ($U_w$; equation (2))[37] below) is then calculated at water depths ($h$ in units of m; 0–200 m at 0.5 m intervals) for the same wave periods and significant wave heights ($H_s$) of 0.5 to 4 m at 0.1 m increments. The resulting maximum water depth (WD; equation (3)) under which the grain size ($D_{90}$) can be transported is then given by the water depth $h$ of the wave-induced near-bed velocity ($U_w$; equation (2)) when the near-bed velocity first exceeds the critical velocity ($U_{cr,w}$; equation (1)) required to initiate transport of the grain size[38].

$$U_{cr,w} = 0.24\left[\left(\frac{\rho_s}{\rho_w}-1\right)g\right]^{0.66} D_{90}^{0.33} T_p^{0.43} \tag{1}$$

$$U_w = \frac{\pi H_s}{T_p \sinh\left(\frac{2\pi h}{L}\right)} \tag{2}$$

$$WD(D_{90}) = h(U_w) \text{ when } U_w > U_{cr,w} \tag{3}$$

$\rho_s$, density of sediment (nominally 2,650 kg m$^{-3}$); $\rho_w$, density of water (1,025 kg m$^{-3}$); $g$, gravity (9.8 m s$^{-2}$); $L$, wave length in m.

Grain size $D_{90}$ is related to the sum of sediment volume percentage >63 μm ($\sum V_{>63}$; percentage sand) to define total bed load transport (Extended Data Fig. 1c). A linear equation for $D_{90}$ is given by equation (4) below ($R^2 = 0.95$); the error is defined as the mean deviation ($\mu X_{Dx}$) of the observations ($x_i$) from the model ($m(x)$; equation (5).

$$\sum V_{>63} = 0.4117 D_{90} - 15.695 \tag{4}$$

$$\mu X_{Dx} = |x_i - m(x)| \tag{5}$$

The accuracy of this method is assessed against three modern wave-graded shelf transects[36,39,40] (Supplementary Figs. 3 and 4) with measured grain size and known wave climates, respectively[41,42], including the modern Whanganui shelf (40.028° S, 173.273° E)[43]. The three resulting profiles of water depth (WD) with $\sum V_{>63}$ for the modern wave climates are shown in Extended Data Fig. 1a, together with the corresponding measured samples of the modern transects. The modern Whanganui wave climate is considered a suitable analogue for the Pliocene owing to the similarities between the open embayment of the Whanganui shelf today and the reconstructed Pliocene shelf and shoreline (Fig. 1; Extended Data Fig. 2). Measurements from the Whanganui shelf are only available for water depths shallower than 80 m (ref. [40]), however confidence in the application of the method to greater depths is given by the reproducibility of the model at depth for the three varying wave climates (Extended Data Fig. 1a).

The error of the model is calculated as the mean percentage of deviation ($\mu X_{Dy}$) of the observations ($y$) for each transect from the respective model ($m(y)$; equation (6)). The summed error of the water depth predicted for $\sum V_{>63}$ for a given wave climate scenario is the standard deviation ($\sigma_n$; equation (7) below) of the range of errors provided by equations (5) and (6). Where $\mu^{max}$ and $\mu^{min}$ are the mean ± the respective error from equations (5) and (6).

$$\mu Y_{Dy} = |y - m(y)| \tag{6}$$

$$\sigma_n, \text{where } n = \begin{bmatrix} \mu_{Dx}^{max} \\ \mu_{Dx} \\ \mu_{Dx}^{min} \end{bmatrix} \cdot \begin{bmatrix} \mu_{Dy}^{max} & \mu_{Dy} & \mu_{Dy}^{min} \end{bmatrix} \tag{7}$$

The WD–$\sum V_{>63}$ model profile for the modern Whanganui shelf (Extended Data Fig. 1a) is used as an analogue to determine water depth from sediment grain size distributions of the Siberia-1 core and Rangitikei River outcrop samples, as outlined above[38]. Wave climate parameters used ($H_s = 2.2$ m, $T_p = 20$ s) are representative of effective sediment transport[44] during extreme storm wave climate on the modern shelf[43], which is time averaged by bioturbation in the geological samples.

To assess the affect of variable wave climates on the reconstruction of relative water depth changes (as opposed to absolute water depths), the derivatives of the water depth–grain size models are calculated for a range of wave heights. Significant wave heights of 2.0–2.5 m occur for the range of the mid-Pliocene interglacial westerly wind changes (for 8.5–9.5 m s$^{-1}$)[45] simulated in PlioMIP global climate models[46], equivalent to ±0.5 m s$^{-1}$ from the present[43]. The derivatives of each model are calculated for a ~30% change in sand (average cycle measured in the stratigraphy) and the difference between the 2.0-m and 2.5-m wave height models result in a maximum of 0.5 m water depth that could be attributed to variable wave climates between glacial and interglacial conditions (Extended Data Fig. 1b). We consider present-day wave climate was equivalent to glacial conditions in the mid-Pliocene based on the relationship between the Holocene and Pliocene glacials in the benthic $\delta^{18}O$ stack[3]. Weaker zonal winds during interglacial phases due to a reduced meridional temperature gradient could imply smaller wave climate than at present during interglacials, but uncertainties in the position of the zonal westerlies[46] in climate models precludes knowing what influence this would have on wave height.

**Frequency analysis.** The $2\pi$ multi-taper method time–frequency analyses were obtained from the 'eha' function in Astrochron[47] (a R package) using default values and plotted as normalized power (maximum in each window is normalized to unity). Window size was 400 kyr for the orbital solution, benthic $\delta^{18}O$ stack and IBRD (Fig. 2b–d), and 200 kyr for the PlioSeaNZ record (Fig. 2e), with a step equal to the sampling interval (2 kyr).

**Backstripping.** Backstripping was undertaken to calculate the sediment and water loading components for the long-term basin subsidence before subtracting this from the high-resolution palaeobathymetry to resolve a RSL record[16,48]. Total

stratigraphic thickness from basement to cessation of sedimentation was determined for the Siberia-1 core site and from a 'pseudo well' for the Rangitikei River Valley outcrop (Supplementary Fig. 5). The outcrop thickness of the Rangitikei River Valley Section was scaled to the thickness of the equivalent formation in the pseudo well. The tectonic subsidence (equation (8)) is first calculated for the two sections at formation resolution assuming no change in RSL ($\Delta RSL = 0$). This result is used as the theoretical subsidence in equation (9), after which the high-resolution palaeo-bathymetric records are included in place of the low-resolution data to calculate the Milankovitch-scale RSL record.

$$T_{sub} = S^* \left( \frac{\rho_m - \rho_s}{\rho_m - \rho_w} \right) + WD + \Delta RSL \left( \frac{\rho_m}{\rho_m - \rho_w} \right). \qquad (8)$$

$$\Delta RSL = \left[ S^* \left( \frac{\rho_m - \rho_s}{\rho_m - \rho_w} \right) + WD - T_{sub} \right] \bigg/ \left( \frac{\rho_m - \rho_w}{\rho_m} \right) \qquad (9)$$

$T_{sub}$, tectonic subsidence (m); $S^*$, decompacted sediment thickness (m); $\rho_m$, mantle density (3.18 g cm$^{-3}$); $\rho_s$, density of sediment (lithology dependent; g cm$^{-3}$); $\rho_w$, density of sea water (1.024 g cm$^{-3}$); $\Delta SL$, change in RSL (m); WD, water depth of depositional environment (m).

RSL calculated this way is not registered to a fixed datum (for example, modern sea level) but does provide an estimate of the amplitude of sea-level cycles (Fig. 3). The mean amplitude for the 54 glaciation and deglaciation events and associated uncertainty ($\sigma_a$; equation (10) below) is determined from the range of amplitudes for the individual glacial ($\mu_G$) and interglacial ($\mu_{IG}$) values including error maxima ($\mu^{max}$) and minima ($\mu^{min}$)[38]. Ages are reported for each event as the mid-range, with uncertainty defined as maximum $\pm5$ kyr resolution of the magnetostratigraphy reversals. This takes into account potential variations in sedimentation rate, climate lag in the astronomically tuned geomagnetic polarity timescale[17] and polarity lock-in time.

$$RSL \pm \sigma_a, \text{ where } a = \begin{bmatrix} \mu_G^{max} \\ \mu_G \\ \mu_G^{min} \end{bmatrix} - \begin{bmatrix} \mu_{IG}^{max} & \mu_G & \mu_{IG}^{min} \end{bmatrix} \qquad (10)$$

**Glacial- and hydro-isostatic adjustment modelling.** The contribution of GIA to local RSL changes is evaluated by solving the gravitationally self-consistent sea-level equation[49]. The SLE incorporates all the GIA feedbacks and yields the space- and time-dependent RSL changes that accompany and follow continental (that is, land-based) ice-sheet thickness variations[50,51]. In this paper we use SELEN, a Fortran 90 program[50] that solves the sea-level equation by means of the pseudo-spectral approach[52]. The latter implies that all the relevant quantities are transformed into complex spherical harmonics up to a maximum degree that we limit at 256. This is combined to a spatial discretization that is based on the icosahedral pixelizations of the sphere and that results in hexagonal elements of ~0.3° of radius. Accordingly, the sea-level equation solution consists of spatio-temporal convolutions where ice-sheet thickness variations are coupled to solid Earth responses and propagated through time to account for the time-dependent viscous relaxation of the mantle. At the core of the sea-level equation formalism is the concept that, at any time $t$ since the beginning of the ice-sheet model chronology, the RSL changes of each point of the Earth's surface stem from the solid Earth and geoid deformations induced by all the ice- and water-loading variations that have occurred since the initial time $t_0$. SELEN requires two main inputs: (i) an ice-sheet model, which describes the ice thickness variation in space and time, and (ii) a rheological model, which describes the solid Earth and gravitational response to ice- and equivalent water-load redistribution.

*Solid Earth model and mantle viscosity profiles.* The solid Earth is assumed to be spherically symmetric, radially stratified, self-gravitating, rotating and deformable, but not compressible[30]. For our reference solid Earth model, we assume a 90-km-thick elastic lithosphere and a Maxwell viscoelastic mantle. We divide the latter into four layers: lower mantle, lower transition zone, upper transition zone and upper mantle. The core of the Earth is considered inviscid. We perform a volume-average of the relevant solid Earth parameters as a function of depth. Accordingly, we use PREM[53] and the VM2 profile for viscosity[54] (Extended Data Table 2). We also investigate the role of mantle viscosity profile and of lithosphere thickness and consider three alternative Earth models.

*Ice-sheet models.* The ANICE-SELEN model[19,55] is used for an LGM to present-day configuration for both GIS and AIS. This allows consistent comparison between all scenarios tested, which include various combinations of between 5–30 m SLE contributions from the Northern and Southern hemispheres. Regional variation in mass contributions from Antarctica are not explored here, because previous studies of various Antarctic ice geometries (LGM to present[56]; deglaciation from

present[10,57,58]) have shown RSL at New Zealand generally approximates ESL for a range of plausible EAIS and WAIS dominated meltwater sources and is within the uncertainties of the PlioSeaNZ record. The only Antarctic scenario that would produce substantial deviation from ESL at New Zealand is mass loss solely from the EAIS, which is unsupported by geological evidence and model simulations[23,29,55,59–61]. This would result in a smaller than eustatic signal at New Zealand particularly in the South Island[62,63]. Such behaviour is usually predicted in South America for a collapse of the WAIS[58,64]. Hence, we have adopted a well-tested and regularly used LGM to present-day configuration with uniform melt distribution. *Ice-sheet scenarios.* We consider first three main eustatic scenarios (15, 20 and 25 m of ESL) and for each scenario we test three different GIS and AIS combinations (AIS only, AIS and GIS in phase, AIS and GIS in anti-phase) for 20-kyr cyclicity, with GIS contributing 5 m and the remainder sourced from AIS (Extended Data Fig. 4a–c). Second, we consider variable periodicities for NHIS and AIS, with larger contribution from the NHIS (GIS, North America and Eurasia). We keep the AIS SLE contribution fixed at 10 m with a 20-kyr periodicity and for the NHIS we assume a 41-kyr periodicity with three different SLE scenarios: 10 m, 20 m and 30 m (Extended Data Fig. 4d–f). We also investigate the effects of extended glacials and interglacials (ice thickness is fixed for 5 kyr rather than linear melting) that show a slightly larger RSL fluctuation because the viscous relaxation of the Earth (for example, ocean syphoning) has more time to respond (Extended Data Fig. 5a–c). The ICE-5G AIS configuration[54] was used to test the effect of the distribution of ice (with the same volume variation), which show differences of up to 0.2 m and are considered secondary to the uncertainty of the RSL of ~$\pm5$ m (Extended Data Fig. 5d). Four Earth models with variable mantle radial viscosity structure are also tested for a 15-m ESL change for the three GIS and AIS contributions, which show RSL variability normalized to ESL of no more than 7% (Extended Data Table 2; Extended Data Fig. 6). Finally, an instantaneous melting event of 15 m, from AIS only, after 10 kyr of viscous relaxation displays RSL at Whanganui within 10% of ESL (Extended Data Fig. 7). All the results (Fig. 4, Extended Data Figs. 4–7) show that the predicted RSL curve at Whanganui is within 10% of eustatic, and therefore can be used to constrain the global ice-volume fluctuations.

## Data availability
The PlioSeaNZ RSL curve and relative amplitudes displayed in Figs. 2 and 3 are available from https://doi.org/10.1594/PANGAEA.902701.

## Code availability
The code for the palaeobathymetry-grain size method is available from https://doi.org/10.1594/PANGAEA.902701.

31. Journeaux, T. D., Kamp, P. J. J. & Naish, T. R. Middle Pliocene cyclothems, Mangaweka Region, Wanganui Basin, New Zealand: a lithostratigraphic framework. *NZ J. Geol. Geophys.* **39**, 135–149 (1996).
32. Ogg, J. G. Geomagnetic polarity time scale. In *The Geologic Time Scale 2012* (eds Gradstein, F. M., Ogg, J. G., Schmitz, M. & Ogg, G.) 85–113 (Elsevier, 2012).
33. Turner, G. M. et al. A coherent middle Pliocene magnetostratigraphy, Wanganui Basin, New Zealand. *J. R. Soc. NZ* **35**, 197–227 (2005).
34. Swift, D. J. Quaternary shelves and the return to grade. *Mar. Geol.* **8**, 5–30 (1970).
35. Wright, J., Colling, A. & Park, D. (eds) *Waves, Tides, and Shallow-Water Processes* Vol. 4 (Gulf Professional Publishing, 1999).
36. Dunbar, G. B. & Barrett, P. J. Estimating paleobathymetry of wave-graded continental shelves from sediment texture. *Sedimentology* **52**, 253–269 (2005).
37. Komar, P. D. & Miller, M. C. On the comparison between the threshold of sediment motion under waves and unidirectional currents with a discussion of the practical evaluation of the threshold: Reply. *J. Sedim. Res.* **45**, 362–367 (1975).
38. Grant, G. R. et al. *A Pliocene relative sea level record from New Zealand calculated from grain size.* https://doi.pangaea.de/10.1594/PANGAEA.902701 (PANGAEA, 2019).
39. Chin, J. L. *Late Quaternary Coastal Sedimentation and Depositional History, South-Central Monterey Bay, California.* Ph.D. thesis, San Jose State Univ. (1984).
40. Beaumont, J., Anderson, T. J. & MacDiarmid, A. B. *Benthic Flora and Fauna of the Patea Shoals Region, South Taranaki Bight.* NIWA Client Report No. WLG2012-55 (NIWA, 2013).
41. Hume, T., Gorman, R., Green, M. & MacDonald, I. *Coastal Stability in the South Taranaki Bight – Phase 2: Potential Effects of Offshore Sand Extraction on Physical Drivers and Coastal Stability.* NIWA Client Report No. HAM2013-082 (NIWA, 2013).
42. Scripps Institution of Oceanography. *CDIP: Coastal Data Information Program.* http://cdip.ucsd.edu/themes/cdip?d2=p70&u3=dt:201101:p_id:p70:ibf:1:mode:all:s:156:st:1:t:data (2018).
43. MetOcean View. *MetOcean View Hindcast.* https://hindcast.metoceanview.com/ (2017).
44. McCave, I. N. Wave effectiveness at the sea bed and its relationship to bed-forms and deposition of mud. *J. Sedim. Res.* **41**, 89–96 (1971).
45. Coastal Engineering Research Centre. *Shore Protection Manual* Vols I and II (US Army Corps of Engineers, Washington DC, 1984).

46. Li, X. et al. Mid-Pliocene westerlies from PlioMIP simulations. *Adv. Atmos. Sci.* **32**, 909–923 (2015).
47. Meyers, S. R. *Astrochron: An R Package for Astrochronology.* https://cran.r-project.org/package=astrochron (2014).
48. Kominz, M. A. Late Cretaceous to Miocene sea-level estimates from the New Jersey and Delaware coastal plain coreholes: an error analysis. *Basin Res.* **20**, 211–226 (2008).
49. Farrell, W. E. & Clark, J. A. On postglacial sea-level. *Geophys. J. R. Astron. Soc.* **46**, 647–667 (1976).
50. Spada, G. & Stocchi, P. SELEN: A Fortran 90 program for solving the "sea-level equation". *Comput. Geosci.* **33**, 538–562 (2007).
51. Stocchi, P. et al. MIS 5e relative sea-level changes in the Mediterranean Sea: contribution of isostatic disequilibrium. *Quat. Sci. Rev.* **185**, 122–134 (2018).
52. Mitrovica, J. X. & Peltier, W. R. On postglacial geoid subsidence over the equatorial oceans. *J. Geophys. Res. B* **96**, 20053–20071 (1991).
53. Dziewonski, A. M. & Anderson, D. L. Preliminary reference Earth model. *Phys. Earth Planet. Inter.* **25**, 297–356 (1981).
54. Peltier, W. R. Global glacial isostasy and the surface of the ice-age Earth: the ICE-5G (VM2) model and GRACE. *Annu. Rev. Earth Planet. Sci.* **32**, 111–149 (2004).
55. de Boer, B., Stocchi, P. & Van De Wal, R. A fully coupled 3-D ice-sheet-sea-level model: algorithm and applications. *Geosci. Model Dev.* **7**, 2141–2156 (2014).
56. Milne, G. A. & Mitrovica, J. X. Searching for eustasy in deglacial sea-level histories. *Quat. Sci. Rev.* **27**, 2292–2302 (2008).
57. Milne, G. A., Gehrels, W. R., Hughes, C. W. & Tamisiea, M. E. Identifying the causes of sea-level change. *Nat. Geosci.* **2**, 471–478 (2009).
58. Mitrovica, J. X. et al. On the robustness of predictions of sea level fingerprints. *Geophys. J. Int.* **187**, 729–742 (2011).
59. Yamane, M. et al. Exposure age and ice-sheet model constraints on Pliocene East Antarctic ice sheet dynamics. *Nat. Commun.* **6**, 7016 (2015).
60. Dolan, A. M., de Boer, B., Bernales, J., Hill, D. J. & Haywood, A. M. High climate model dependency of Pliocene Antarctic ice-sheet predictions. *Nat. Commun.* **9**, 2799 (2018).
61. Shakun, J. D. et al. Minimal East Antarctic Ice Sheet retreat onto land during the past eight million years. *Nature* **558**, 284 (2018).
62. Hay, C. et al. The sea-level fingerprints of ice-sheet collapse during interglacial periods. *Quat. Sci. Rev.* **87**, 60–69 (2014).
63. Kopp, R. E. et al. Temperature-driven global sea-level variability in the common era. *Proc. Natl Acad. Sci. USA* **113**, E1434–E1441 (2016); correction. **113**, E5694–E5696 (2016).
64. Bamber, J. L., Riva, R. E., Vermeersen, B. L. & LeBrocq, A. M. Reassessment of the potential sea-level rise from a collapse of the West Antarctic Ice Sheet. *Science* **324**, 901–903 (2009).

**Author contributions** T.R.N. and G.R.G. designed the project. G.R.G. measured and analysed the data. G.R.G., T.R.N. and G.B.D. interpreted the results. P.S., M.A.K., P.J.J.K. and C.A.T. contributed to modelling and supporting datasets. R.A.M., R.H.L. and M.O.P. assisted in interpretation of the data. All authors contributed to drafting of the manuscript.
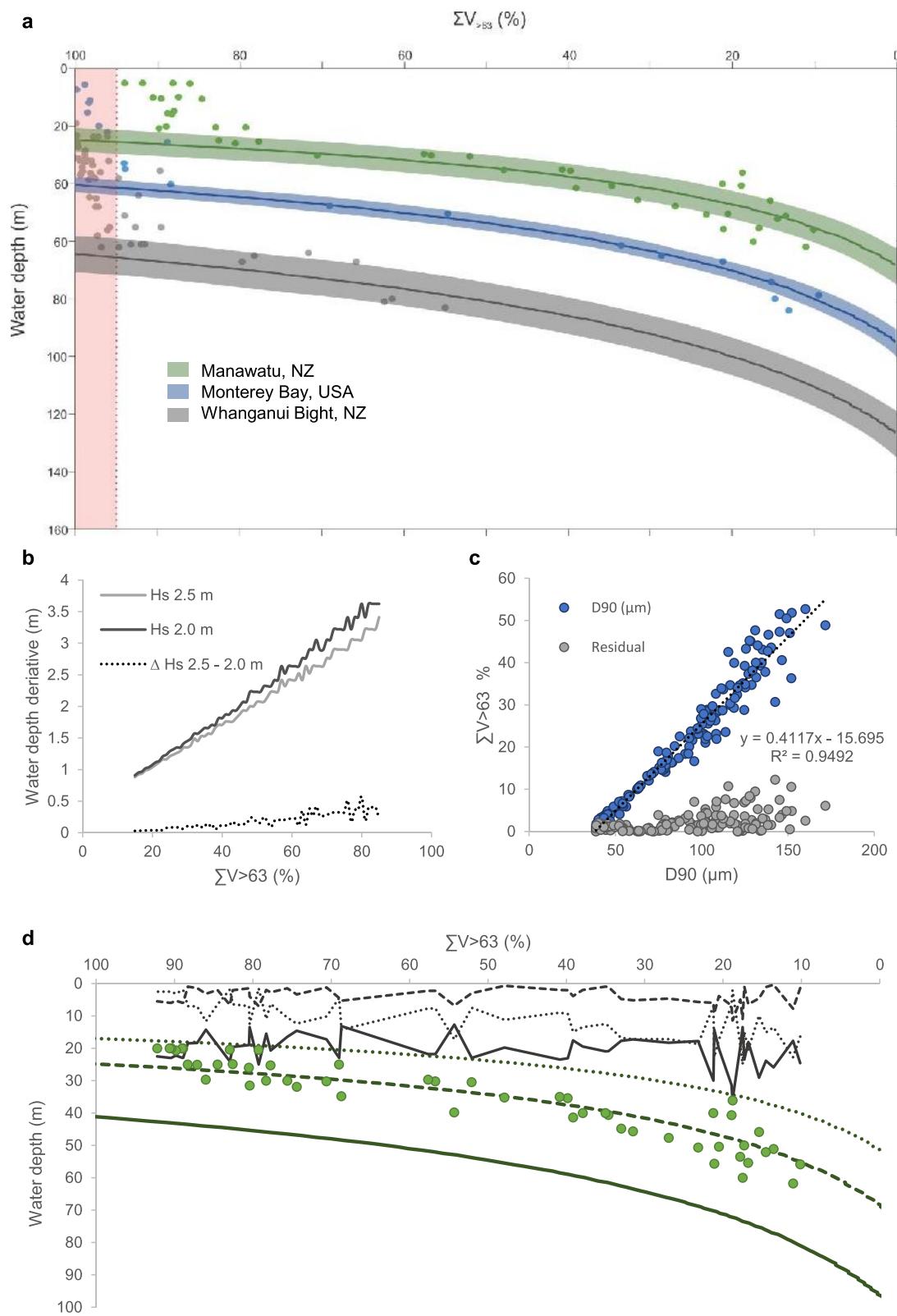
**Extended Data Fig. 1 |** See next page for caption.

**Extended Data Fig. 1 | Modern analogue of the grain size–water depth relation and model calculations. a**, Observations (dots) and model values (shaded bands represent the maximum and minimum ranges from the average bold lines) for sand ($\sum V_{>63}$) and water depth for three different modern shelf transects (Manawatu, NZ, green[36]; Monterey Bay, USA, blue[39]; Whanganui Bight, NZ, grey[40]). Wave parameters used are as follows: for Manawatu, $H_s = 1.2$ m and $T_p = 20$ s (ref. [41]); for Monterey, $H_s = 1.8$ m and $T_p = 20$ s (ref. [42]); and for Whanganui Bight, $H_s = 2.2$ m and $T_p = 20$ s (ref. [43]). See Methods for nomenclature. Model error is described by equation (8). The red shaded band for $\sum V_{>63} = 95\%–100\%$ represents the limit of the method, where all water depths contain 100% $\sum V_{>63}$. The modern Whanganui Bight is selected as the most appropriate modern analogue to determine water depth from $\sum V_{>63}$ recorded in both core and outcrop in this study. **b**, Derivatives of water depth–grain size models, for an average sediment cycle amplitude of 30% $\sum V_{>63}$, for peak wave period $T_p = 20$ s and significant wave height $H_s = 2.0$ m (dark grey) and $H_s = 2.5$ m (light grey) and the difference (dashed light grey). **c**, Calibration of $\sum V_{>63}$ from maximum grain size in distribution and measured $D_{90}$ from core samples (blue circles) described by a linear relationship (dotted dark blue line; equation (5)) and the deviation (grey) of the model from observations. **d**, Calibration of peak wave period ($T_p$) exponent for the critical required velocity ($U_{cr,w}$; equation (1))[14]. Observations for Manawatu (most extensively sampled; green circles) are used for comparison between peak wave period exponents 0.33 (solid dark green line), 0.43 (dashed dark green line) and 0.5 (dotted light green line), with the deviations between models and observations shown by the respective patterned thinner black lines.
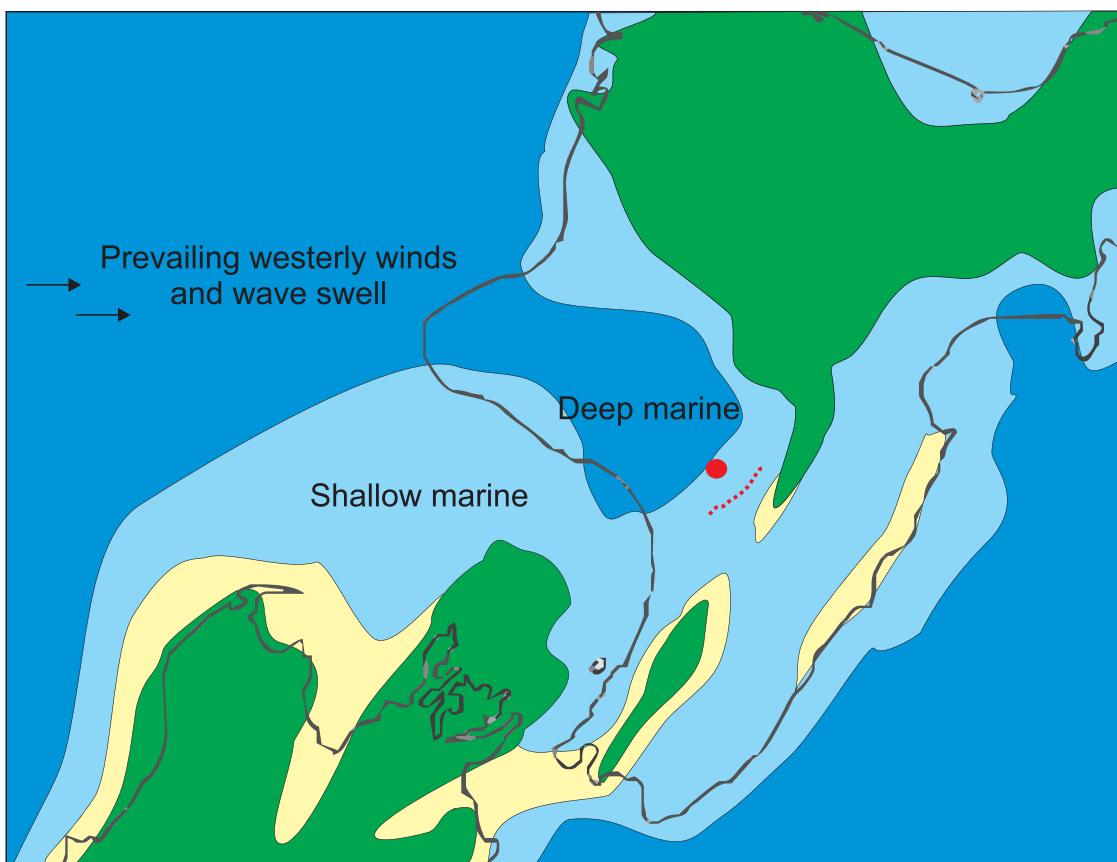
**Extended Data Fig. 2 | Pliocene palaeogeography of the Whanganui Basin.** The figure shows a semi-enclosed broad embayment open to the dominant westerly wind, with an arcuate shoreline and a westward-deepening shelf[15]. The location of the Siberia-1 core (red circle) and the Rangitikei River Section (dotted red line) are noted.
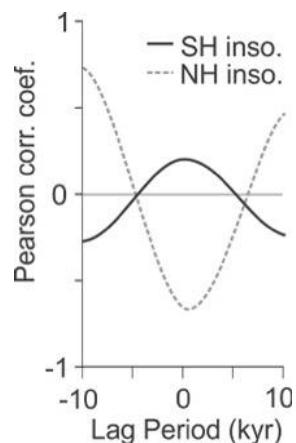
**Extended Data Fig. 3 | Temporal relationship between PlioSeaNZ RSL and latitudinal insolation.** Southern and Northern hemisphere high-latitude summer insolation[18] (65° S, 1 January, solid line; 65° N, 1 July, dashed line) Pearson correlation coefficient with the PlioSeaNZ record between 3.3 and 3.0 Ma, using the 'slideCor' function from the R – Astrochron package[47]. Here a 0-kyr lag period denotes no temporal shift in the untuned PlioSeaNZ age model, and ±10-kyr lag periods signify correlation with a positive or negative shift of the PlioSeaNZ age model with respect to the astronomical record.
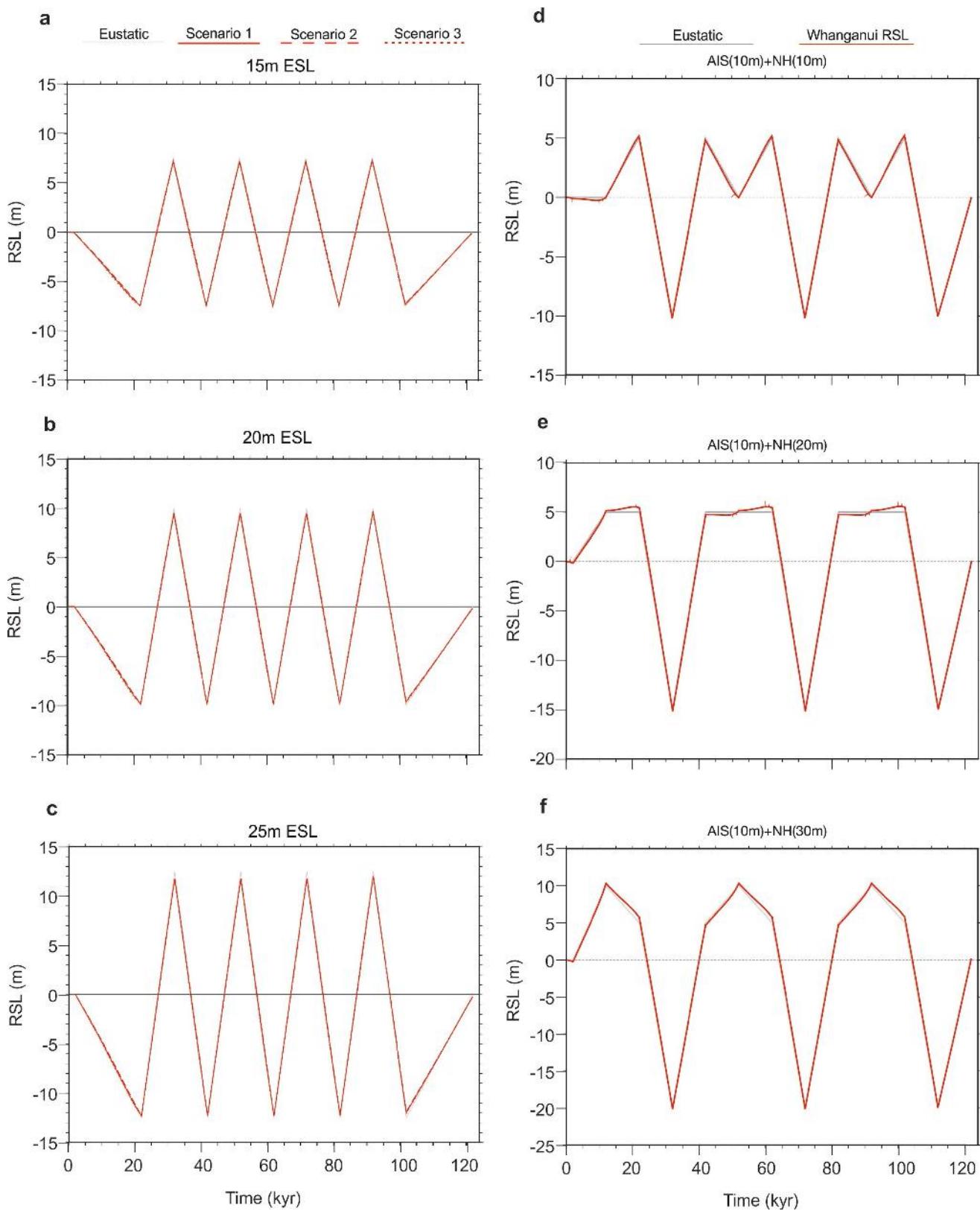
**Extended Data Fig. 4 | Assessment of RSL predicted at Whanganui, New Zealand, for pre-determined ESL scenarios. a–c**, RSL calculated for 20-kyr glacial–interglacial polar ice-sheet variability for three values of ESL (**a**, 15 m; **b**, 20 m; and **c**, 25 m) and three scenarios of polar ice-sheet contribution. Scenario 1 represents an Antarctic-only contribution, scenario 2 represents a Greenland Ice Sheet (GIS) contribution (of 5 m) in phase with a 15-m Antarctic contribution, and scenario 3 has a 30-m AIS contribution in anti-phase with 5 m of GIS accumulation. For each ESL value, all scenarios are indistinguishable as RSL at the Whanganui, New Zealand, site. **d–f**, RSL calculated for 20-kyr Antarctic variability and 40-kyr Northern Hemisphere variability, with 10 m from AIS and three different contributions from Northern Hemisphere ice sheets (NHIS); **d**, 10 m; **e**, 20 m; and **f**, 30 m.
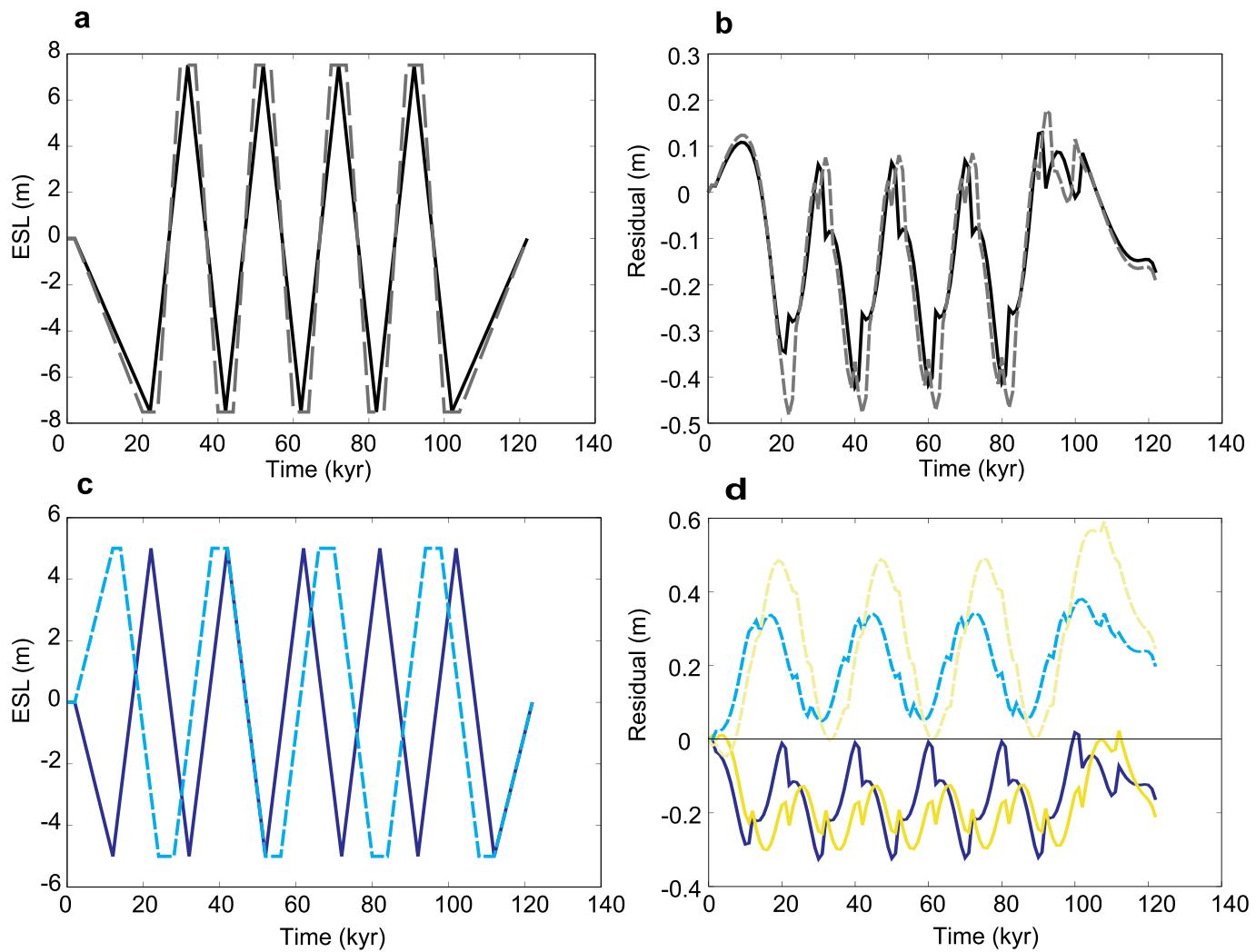
**Extended Data Fig. 5 | Sensitivity analysis of extended 5-kyr glacials and interglacials.** Modelled RSL at Whanganui, New Zealand, for comparison of symmetrical glacial–interglacial cyclicity (bold lines) and extended glacials and interglacials (dashed lines) using the ANICE-SELEN ice-sheet model. **a**, Modelled RSL for a 15-m ESL fluctuation as a symmetrical waveform (bold black line) and with extended glacials and interglacials (dashed grey line); **b**, the residuals of RSL curves from **a** with respect to ESL (symmetrical, bold black; and extended, dashed grey). **c**, As **a** but for a 10-m ESL fluctuation of a linear 20-kyr chronology (bold dark blue line) and a longer period cyclicity from cumulative extended glacial and interglacials (dashed blue line); **d**, as **b** showing the residuals from **c** but repeated using the ICE-5G model[54] (symmetrical bold, yellow; and extended, dashed yellow). The differences between the ANICE-SELEN and ICE-5G models in **d** are evident but are of the order of tens of centimetres. Interestingly, longer periodicity and extended glacials/interglacials yield larger RSL excursions with respect to ESL (positive values).
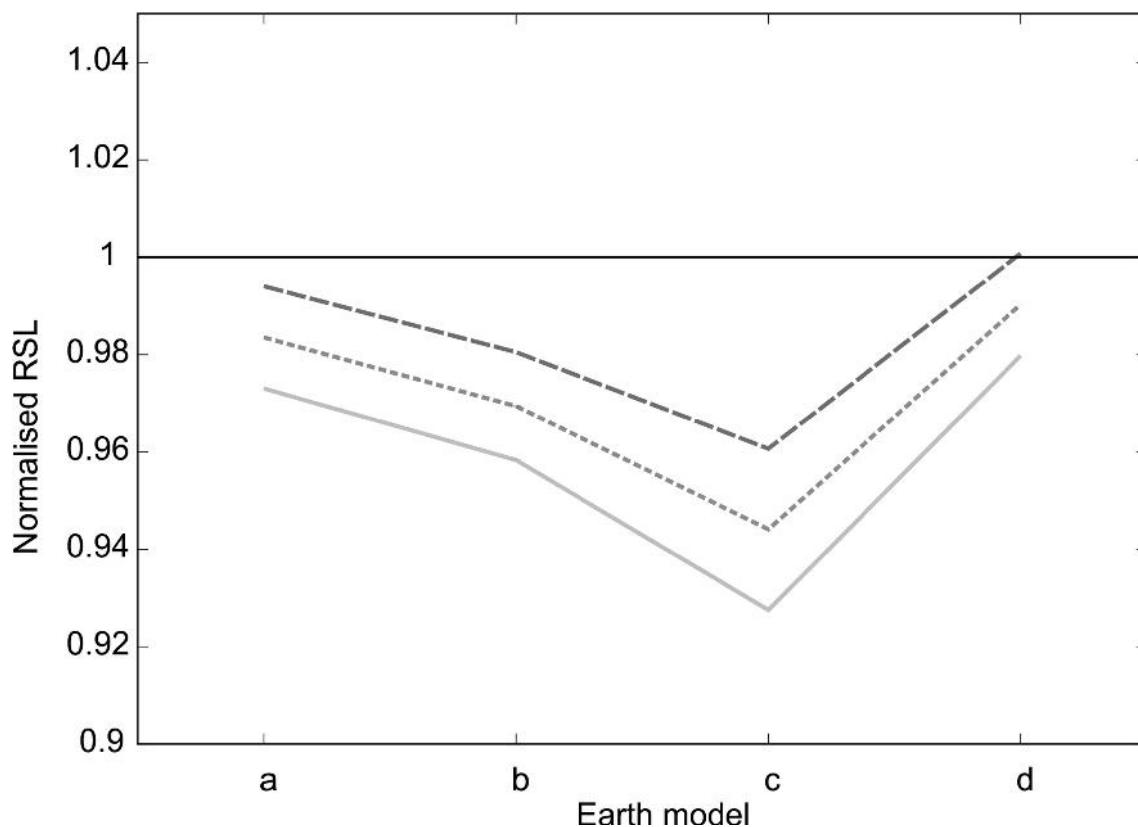
**Extended Data Fig. 6 | Predicted RSL rise at Whanganui for different mantle viscosity profiles.** Shown is the predicted RSL rise for a 15-m ESL change, after 10 kyr of linear melting between glacial and interglacial for scenarios described in Extended Data Fig 4 (namely, scenarios 1 (short-dashed line), 2 (dashed line) and 3 (solid grey line)), for the four Earth models (a, b, c, d; Extended Data Table 2). Higher contrast between lower and upper mantle results in lower values for the predicted RSL rise. A thicker lithosphere (120 km) results in a slightly higher than eustatic peak for scenario 1.
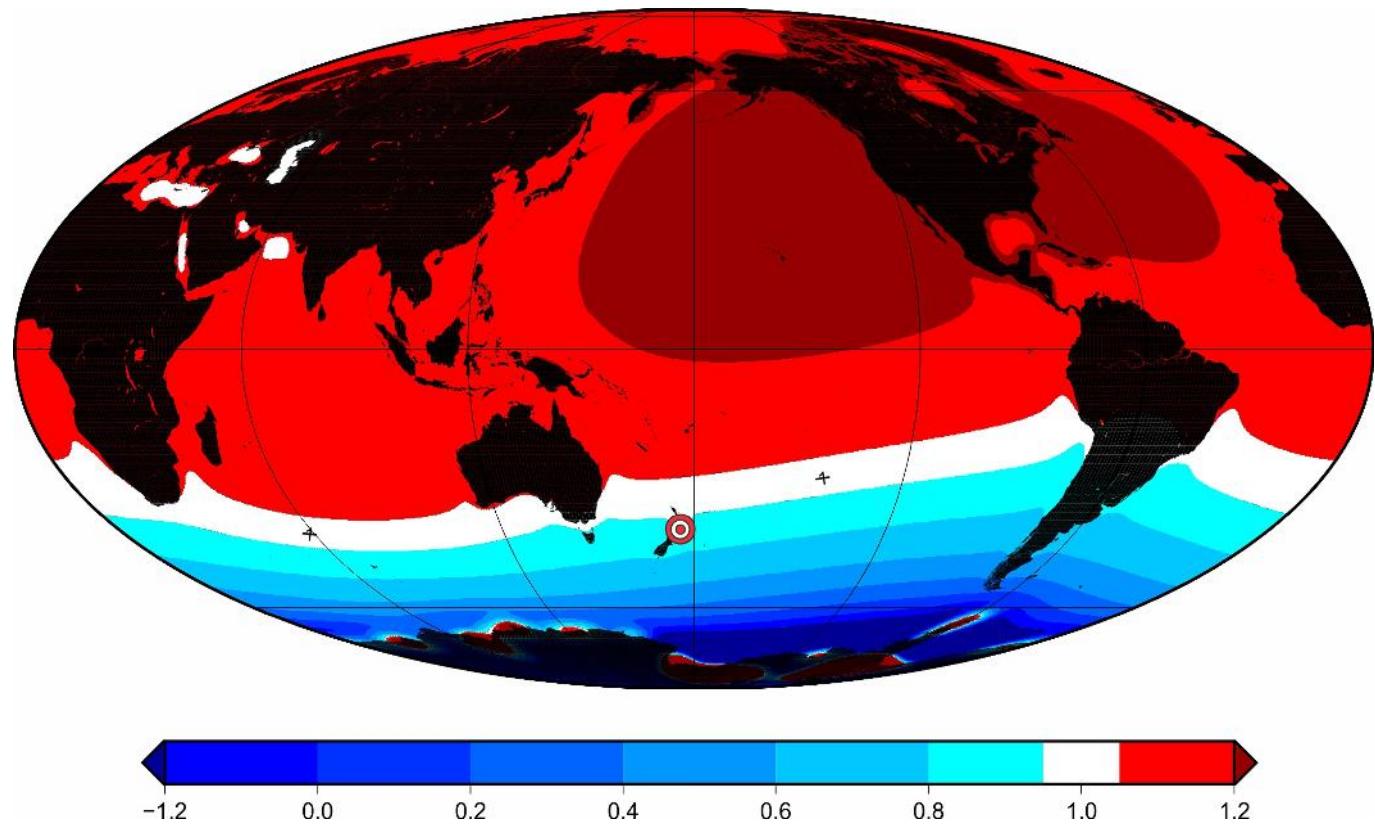
**Extended Data Fig. 7 | Calculated global RSL change produced by instantaneous ice-sheet melting of 15-m ESL.** Shown is RSL (normalized with respect to ESL) according to scenario 1 (AIS only) after 10 kyr of viscous relaxation (mantle viscosity profile a; Extended Data Table 2) following an instantaneous melting. The white band denotes RSL from 0.8 to 1.2 of the ESL signal. The GIA-driven RSL fingerprints are more evident if compared to Fig. 4a (10-kyr-long linear melting). The Whanganui site is highlighted by the red and white bullseye.

**Extended Data Table 1 | Age tie points and associated linear sedimentation rates**

| Section | Depth (m) | Thickness (m) | Age (Ma) | Linear sedimentation rate (m.ky⁻¹) |
|---|---|---|---|---|
| Rangitikei River Section | 785 | | 2.581^ | |
| | | 410 | | 0.91 |
| | 375 | | 3.032^ | |
| Siberia-1 | (exhumed surface) | | 2.581^ | |
| | | 763 | | 1.69 |
| | 82.54 | | 3.032* | |
| | | 64.46 | | 0.77 |
| | 147 | | 3.116* | |
| | | 126.98 | | 1.4 |
| | 273.98 | | 3.207* | |
| | | 66.02 | | 0.71 |
| | 340 | | 3.300‡ | |

Data are shown for the Siberia-1 core and the Rangitikei River Section outcrop. The tie points are based on palaeomagnetic boundaries recorded in the outcrop[33] (^) and core[6] (*), with the exception of a tie point (‡) at the base of the Siberia-1 core, which is correlated to the M2 glaciation in the regional chronostratigraphic framework[6].

**Extended Data Table 2 | Solid Earth model parameters used in GIA computations**

| Number of interfaces | Radius (km) | Density (kg.m⁻³) | Shear moduli (x 10¹⁰ Pa) | Viscosity[a] (x 10²¹ Pa·s) | Viscosity[b] (x 10²¹ Pa·s) | Viscosity[c] (x 10²¹ Pa·s) | Viscosity[d] (x 10²¹ Pa·s) |
|---|---|---|---|---|---|---|---|
| 1 | 3480 | 10931 | 0.0 | 0.00 | 0.00 | 0.00 | 0.00 |
| 2 | 5701 | 4877 | 22.0 | 2.53 | 5.00 | 10.00 | 2.53 |
| 3 | 5971 | 3857 | 10.6 | 0.46 | 0.46 | 0.46 | 0.46 |
| 4 | 6151 | 3475 | 7.6 | 0.44 | 0.44 | 0.44 | 0.44 |
| 5 | 6281* | 3370 | 6.6 | 0.67 | 0.67 | 0.67 | 0.67 |
| 6 | 6371 | 3192 | 6.0 | 0.00 | 0.00 | 0.00 | 0.00 |

The actual Maxwell viscoelastic mantle is represented by 'Number of interfaces' 2–5 (4 layers). The complete mantle viscosity profile 'a' used here corresponds to the volume average of the original VM2 profile[54]. Mantle viscosity profiles 'b' and 'c' are characterized by higher viscosity values for the lower mantle. Profile 'd' has the same values as profile 'a' but with a thicker elastic lithosphere (120 km; *radius of 6,251 km).

# LETTER

# Diversity decoupled from ecosystem function and resilience during mass extinction recovery

Sarah A. Alvarez[1,2,6]*, Samantha J. Gibbs[3]*, Paul R. Bown[2], Hojung Kim[2], Rosie M. Sheward[4] & Andy Ridgwell[5]

The Chicxulub bolide impact 66 million years ago drove the near-instantaneous collapse of ocean ecosystems. The devastating loss of diversity at the base of ocean food webs probably triggered cascading extinctions across all trophic levels[1–3] and caused severe disruption of the biogeochemical functions of the ocean, and especially disrupted the cycling of carbon between the surface and deep sea[4,5]. The absence of sufficiently detailed biotic data that span the post-extinction interval has limited our understanding of how ecosystem resilience and biochemical function was restored; estimates[6–8] of ecosystem 'recovery' vary from less than 100 years to 10 million years. Here, using a 13-million-year-long nannoplankton time series, we show that post-extinction communities exhibited 1.8 million years of exceptional volatility before a more stable equilibrium-state community emerged that displayed hallmarks of resilience. The transition to this new equilibrium-state community with a broader spectrum of cell sizes coincides with indicators of carbon-cycle restoration and a fully functioning biological pump[9]. These findings suggest a fundamental link between ecosystem recovery and biogeochemical cycling over timescales that are longer than those suggested by proxies of export production[7,8], but far shorter than the return of taxonomic richness[6]. The fact that species richness remained low as both community stability and biological pump efficiency re-emerged suggests that ecological functions rather than the number of species are more important to community resilience and biochemical functions.

The end Cretaceous bolide impact stripped the ocean of diversity and biogeochemical function[1] more abruptly than any other mass extinction event, including the current anthropogenically induced crisis. After extinction of more than 90% of species of calcifying plankton[3], the oceans were repopulated in the immediate aftermath of the impact by aberrant communities dominated by ephemeral species that were atypical in ecology, physiology and cell size[9–11]. Over time, a diverse, biochemically functioning and resilient ecosystem was re-established. This complete re-assembly of the ocean ecosystem provides clues to the essential attributes that underpin stable ecosystems and maintain robust ecological states and function[12,13]. However, assessments of
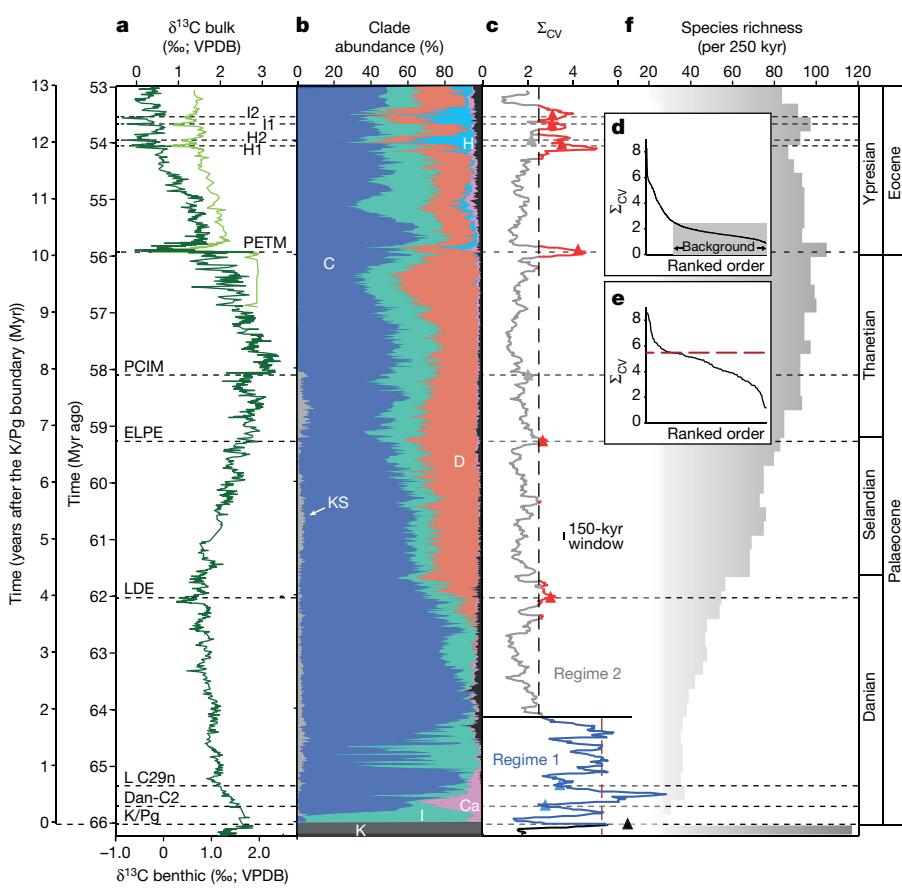


**Fig. 1 | Nannoplankton abundance, variability and diversity records from the latest Cretaceous to early Eocene. a**, Carbon stable isotopes (see references in a previously published study[21]). Dark green, benthic[32]; light green, bulk. Event nomenclature follows references as described in the Methods. **b**, Summary of the abundances of the main nannoplankton clades determined for 981 samples. C, Coccolithales; Ca, calcispheres; D, Discoasterales; H, holococcoliths; I, Isochrysidales; K, Cretaceous; KS, Cretaceous survivors. **c**, $\Sigma_{CV}$ using 150-kyr moving windows (see Methods), separated into the early Danian (regime 1, blue) and the rest of the record (regime 2, grey and red). The Cretaceous to K/Pg data points are shown in black. Triangles (colour-coded by regime, black for the K/Pg window) show values for the named climate events that, for the Eocene hyperthermals, were calculated across the event duration (<150 kyr). The vertical black dashed line in **c** indicates the estimated background level (<2.5 $\Sigma_{CV}$) based on the ranked order inflection point shown in **d** (grey background) and above-background data are highlighted in red in **c**. **d**, Ranked-order plot for all data points. **e**, Ranked-order plot of regime 1 (early Danian) data points. The dark red dashed line indicates an inflection point at the higher end of the $\Sigma_{CV}$ values (shown as a red dashed line in **c**). **f**, Calcareous nannoplankton species richness at 250-kyr resolution (see Methods).

[1]School of Geographical Sciences, University of Bristol, Bristol, UK. [2]Department of Earth Sciences, University College London, London, UK. [3]Ocean and Earth Science, National Oceanography Centre Southampton, University of Southampton, Southampton, UK. [4]Institute for Geoscience, Goethe-Universität Frankfurt, Frankfurt am Main, Germany. [5]Department of Earth Sciences, University of California at Riverside, Riverside, CA, USA. [6]Present address: University of Gibraltar, Europa Point Campus, Gibraltar, Gibraltar. *e-mail: sarah.alvarez@unigib.edu.gi; sxg@noc.soton.ac.uk
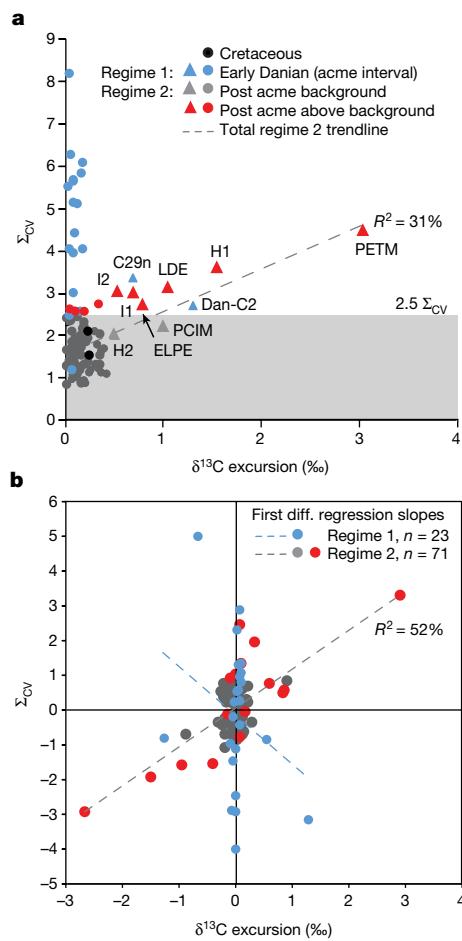
**Fig. 2 | $\Sigma_{CV}$ and magnitude of climate perturbation ($\delta^{13}$C excursion). a**, Raw data of the $\Sigma_{CV}$ and $\delta^{13}$C excursion. **b**, First differences in successive points in the time series of the $\Sigma_{CV}$ and $\delta^{13}$C. The $\Sigma_{CV}$ values are plotted for each named climate event (using the highlighted values in Fig. 1c, shown as triangles in **a**) and every intervening approximately 150 kyr. Data points are separated into Cretaceous (black, $n = 2$; data not included in **b**), early Danian (blue, 66–64.2 Myr ago; regime 1, $n = 23$) and the rest of the record (grey and above background in red; regime 2, $n = 71$; data shown in Fig. 1c, d). The $\Sigma_{CV}$ of regime 1 shows no relationship with climatic perturbation (blue dashed trend line in **b**), in contrast to the $\Sigma_{CV}$ of regime 2 (generalized least squares trend line for all grey plus red data points, $n = 71$; grey dashed line in **a** and **b**), which is significant for both the raw data (**a**) and the first differences (**b**) ($R^2 = 31\%$ and 52%, respectively (see Methods)). The inferred background level of 2.5, based on rank ordering (Fig. 1d), is indicated by the grey box.

when this ecosystem 'recovery' was achieved vary widely in definition and duration. Proxies of export production suggest nearly instantaneous restoration of at least some biogeochemical functionality, in less than 100 years[7,8], while the analysis of the return of species richness to pre-extinction levels suggests that recovery occurred 8–10 million years (Myr) later[6]. Here, we track the post-extinction path to ecosystem restoration by building a high-resolution 13-Myr-long community record of calcareous nannoplankton, the dominant fossil-forming primary producers. Many species of the marine food web leave little or no fossil remains; however, the biomineralized exoskeletons of calcareous nannoplankton provide a remarkable proxy for basal ecosystem health during past environmental change events (see ref. [3] and references therein). Our nannoplankton record bridges the temporal range of current recovery estimates and allows us to target measures of community stability (the level of deviation around the average state; see Methods) and resilience (the ability to resist and recover from perturbation[14]) as they re-emerged. The record from Ocean Drilling Program Site 1209 in the Pacific Ocean (Extended Data Fig. 1 and Methods) has highly

resolved orbital age control (see Methods) and complementary proxy data for environmental change and biogeochemical function. Our plankton data comprise a sample every approximately 13 thousand years (kyr), span 13 Myr and consist of around 700,000 fossil counts, providing a time series of key community parameters including abundance, diversity, taxonomic richness, variance, dissimilarity and body size (see Methods).

Our data and analyses reveal striking temporal trends in the structure and resilience of the nannoplankton community (Fig. 1). Most visually apparent is the differentiation of a highly volatile post-extinction interval of around 1.8 Myr, from a subsequent more-stable background state (Fig. 1e), as can be seen in the summed coefficient of variation metric ($\Sigma_{CV}$). We primarily focus on this metric, which quantifies the level of variance or stability in relative abundances (see Methods); however, as community stability is a multifaceted concept, we also refer to other indices, including community dissimilarity (Bray–Curtis dissimilarity) and diversity (Simpson's index of dominance/evenness). These measures of community structure all point to a state shift around 1.8 Myr after the impact (Fig. 1 and Extended Data Fig. 2) and an early Danian interval characterized by strong fluctuations that are statistically distinct from the rest of the record (Extended Data Fig. 2), which are hereafter referred to as regime 1 (66.0–64.2 Myr ago) and regime 2 (64.1–53.0 Myr ago), respectively. When we compare $\Sigma_{CV}$ with carbon isotope ($\delta^{13}$C) excursion magnitude, which is a proxy for environmental change (Fig. 2 and Methods), the two regimes show markedly different relationships with environmental forcing. The earliest Danian (regime 1) exhibits no relationship between $\Sigma_{CV}$ and $\delta^{13}$C magnitude. Prolonged high-amplitude variance is mainly the statistical effect of a series of ocean-wide abundance peaks (acmes)[10,15,16] (Fig. 3a, b and Extended Data Figs. 3, 4) and occurs alongside very little apparent environmental perturbation (Figs. 1a, 2). During this interval, there were very short-term (much less than 100 kyr) impact-related environmental changes[17,18] (cooling[17] during less than 50 years, and subsequent warming[18] for less than 100 kyr), together with waning Deccan trap volcanism[19] during 600 kyr. Furthermore, only two notable environmental change events have been identified—the lower-C29n and Dan-C2 hyperthermals. All of these changes occurred or ceased well before the interval of high variance comes to an end. Therefore, the disconnect between community metrics and indicators of climate variability suggests that environmental changes were not driving and maintaining the high levels of biotic variability through this 1.8-Myr-long interval. By contrast, above this stratigraphic level (regime 2), $\delta^{13}$C magnitude is a strong predictor of community variance ($R^2 = 52\%$ when using first differences of both variance and $\delta^{13}$C magnitude; Fig. 2) in which most of the data form a 'background' grouping that is punctuated by variance peaks associated with hyperthermal events[20] (highlighted in red in Figs. 1a, c, 2). This indicates that by the late Danian, nannoplankton communities fluctuated around a steady-state condition and demonstrated indicators of resilience[14], such as proportionate responses to environmental perturbation (that is, a significant linear trend between carbon isotope excursion and variance) and rapid recovery after each perturbation event (that is, the return of variance to the background state within less than 200 kyr of the excursion[21]; Fig. 1c).

Notably, the shift to more-stable communities (approximately 64.2 Myr ago; the end of regime 1) also occurs towards the top of the interval during which the biological pump recovered[9] (Fig. 3f). Ocean biogeochemical function was markedly disrupted by the end-Cretaceous mass extinction, most obviously through weakening of the biological pump[2,5,9]. The scale and duration of this productivity reduction is contentious, ranging from scenarios of a lifeless 'Strangelove ocean' to a partially functioning 'living ocean' state[4]; however, the long, multi-million-year delay in restoration of the biological pump is well-established[2,22] and is indicated by both the gradual increase in the vertical carbon isotope gradient to pre-extinction values[9] and the changing community structures of benthic primary-consumer communities (benthic foraminifera)[23]. Carbon isotope gradients finally returned to pre-extinction values around 1.77 Myr after the extinction
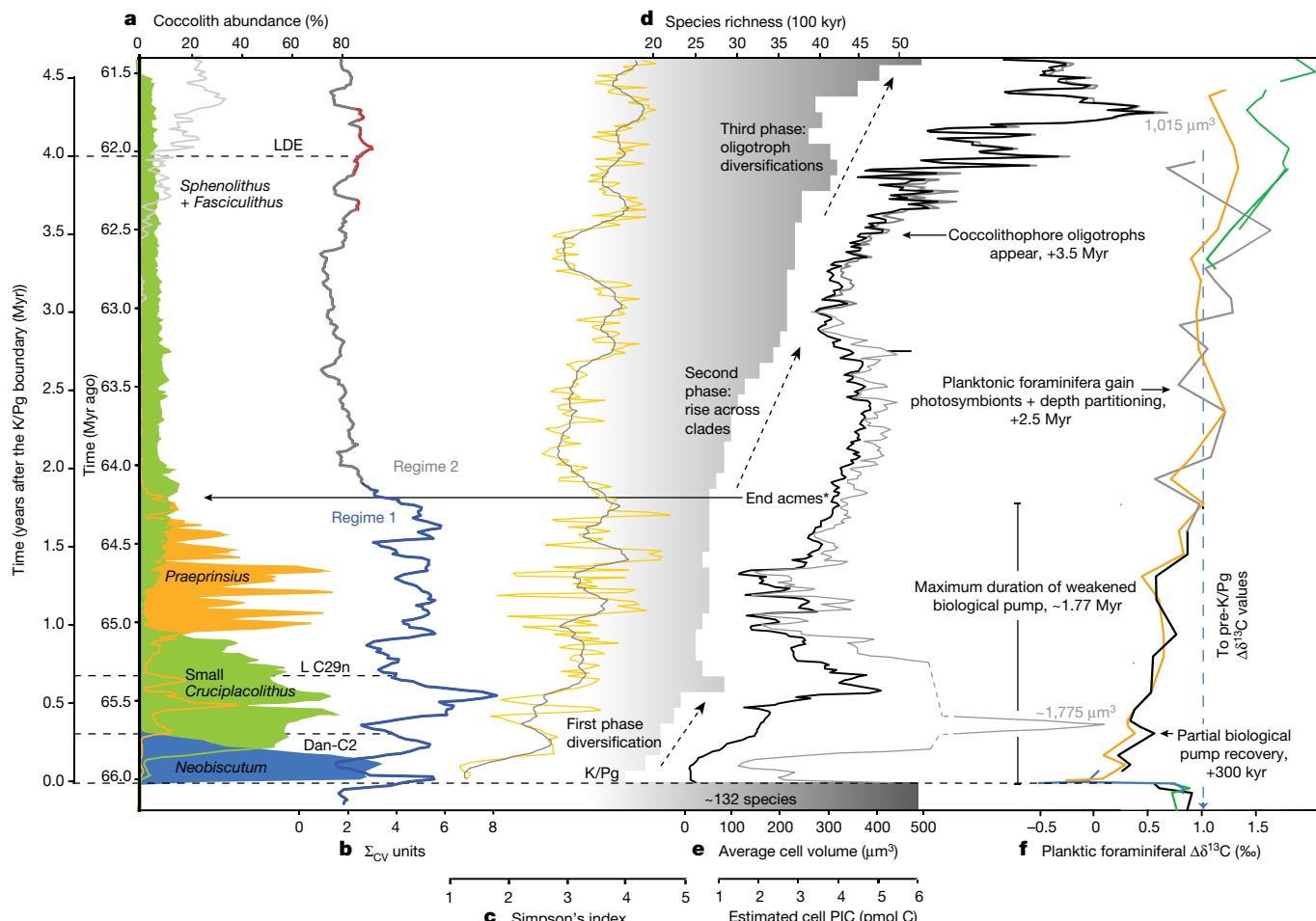
**Fig. 3 | Danian nannoplankton community variance, acme abundances, diversity, cell volume and key milestones. a**, Abundance records of the early Danian coccolithophore acmes and Discoasterales (*Sphenolithus* and *Fasciculithus*). **b**, $\Sigma_{CV}$ (regime 1 in blue, regime 2 in grey and red) and Danian climate events. **c**, Community diversity (Simpson's index); the 150-kyr moving average is shown in grey. **d**, Global species richness resolved at the 100-kyr scale (see Methods). **e**, Estimated average (mean) cell volume of the calcareous nannoplankton (grey), excluding calcispheres (black) (see Methods). The PIC axis indicates the corresponding carbonate content estimated for the calcareous nannoplankton black line based on a linear regression between fossil nannoplankton PIC and cell volume

(see Methods). Cell size maxima at 300 kyr and 4.25 Myr after the K/Pg are shown (grey). Note the break in scale for the lower peak between 525 and 1,450 μm³. Key milestones/observations from this study and from published records are indicated (see Methods). The level for the end acmes is taken as the top of the *Praeprinsius* acme (see asterisk). **f**, Carbon isotope gradient ($\Delta\delta^{13}C$) between Walvis Ridge (Ocean Drilling Program Site 1262) planktonic and benthic foraminifera species[9] (analysed using adjustment option 2)[9], grouped according to ecology. Blue, surface survivors; green, surface symbiotic; orange, thermocline; black, mixed layer; grey, transitional.

event[9], providing an upper limit to the full recovery of the biological pump. This broad concurrence between biological pump restoration and the shift to a more-stable background state in the plankton community (Fig. 3) provides strong evidence for an intrinsic link between biological recovery of the ecosystem and its calibre of biochemical functioning. We can expand our understanding of ecosystem recovery and efficient biological pumping by exploring the roles of the post-extinction taxonomic structure and rapid increases in cell size using high-resolution data on species richness (Fig. 3d and Methods) and reconstructions of the cell volumes of the nannoplankton community (Fig. 3e and Methods).

Mean cell volume of the community and species richness exhibited pulsed patterns through the Danian; both showed rapid increases in the first 0.5 Myr after the mass extinction from initially extremely low species numbers and predominantly very small cells (Fig. 3d, e). Rapid diversification within regime 1 saw the appearance of more than 15 species alongside a peak in cell volume (around 300 kyr after the extinction level) that was dominated by cells of heavily calcified calcareous dinoflagellates. A second phase of increase in cell volume occurred as carbon export gradually returned to pre-extinction values, driven by both diversity and ecology. The relative abundance of existing

large taxa increased (such as *Coccolithus*; Extended Data Fig. 3) and larger new species appeared across all of the emerging clades (Fig. 3d). Observations of modern plankton assemblages indicate that the size structure of phytoplankton communities is an important control on export flux and that nannoplankton mineral ballasting considerably increases the transfer efficiency of carbon[24–26]. The shift to larger cells and ballast biominerals that is evident in our early Danian cell-size record (Fig. 3e) would therefore have contributed to increased carbon export flux, with stable, diverse communities delivering this flux more consistently through space and time, and would have supported greater size diversity in the zooplankton[26]. The role of larger zooplankton and the production of fast-sinking faecal pellets in these evolving export pathways are more difficult to reconstruct owing to poor fossil records. However, an indication of the disruption of higher trophic levels is seen in the early Danian zooplanktonic foraminifera, which show low diversities, acme fluctuations and small body sizes across similar timescales to the recovery of the biological pump[11,27]. Finally, a third phase of increase in cell volume coincides with a major expansion of ecological diversity that is marked by the appearance of the first specialist oligotrophic nannoplankton since the mass extinction (Discoasterales; around 3.5 Myr after the extinction event[28];

Figs. 1b, 3a, d, e and Extended Data Fig. 4) and reintroduction of photosymbiotic strategies in planktonic foraminifera (around 2.5 Myr after the extinction event[9]; Fig. 3f). Diversification then continued, with species richness only reaching pre-extinction levels at 10 Myr after the mass extinction (approximately 56 Myr ago; Fig. 1g).

The scale of the ecosystem collapse associated with the Cretaceous–Palaeogene (K/Pg) event and the protracted recovery of resilience, diversity and biogeochemical function demonstrate the considerable consequences of mass-extinction-level change and the subsequent durability of the ecosystems after restoration. Predictions of a contemporary mass extinction[29] highlight accelerating declines in ecosystem functioning as diversity decreases[30,31]. Here, we show that the reverse also holds as biodiversity recovers after a mass extinction. Early, albeit modest, diversification of taxa and traits (for example, cell size) within the recovering biota re-established the stability of the ecosystem with links to biological functions (such as the biological pump) long before species richness and ecological diversity returned to pre-extinction levels. Rapid biotic turnover and community instability during this reboot increased the probability that biotically forced evolution alighted on organisms that were capable of fulfilling essential functional roles, and this, in turn, facilitated community recovery and the re-emergence of ecosystem stability. Ecosystem stability is therefore not only determined by the numbers of species, but also by the establishment and/or retention of key functional taxa that fulfil vital ecological and/or biogeochemical roles.

## Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-019-1590-8

1. Schulte, P. et al. The Chicxulub asteroid impact and mass extinction at the Cretaceous–Paleogene boundary. *Science* **327**, 1214–1218 (2010).
2. Coxall, H. K., D'Hondt, S. & Zachos, J. C. Pelagic evolution and environmental recovery after the Cretaceous–Paleogene mass extinction. *Geology* **34**, 297–300 (2006).
3. Bown, P. R., Lees, J. A. & Young, J. R. in *Coccolithophores: from Molecular Processes to Global Impact* (eds Thierstein, H. R. & Young, J. R.) 481–508 (Springer, 2004).
4. D'Hondt, S. Consequences of the Cretaceous/Paleogene mass extinction for marine ecosystems. *Annu. Rev. Ecol. Evol. Syst.* **36**, 295–317 (2005).
5. Henehan, M. J., Hull, P. M., Penman, D. E., Rae, J. W. B. & Schmidt, D. N. Biogeochemical significance of pelagic ecosystem function: an end-Cretaceous case study. *Phil. Trans. R. Soc. B* **371**, 20150510 (2016).
6. Kirchner, J. W. & Weil, A. Delayed biological recovery from extinctions throughout the fossil record. *Nature* **404**, 177–180 (2000).
7. Sepúlveda, J., Wendler, J. E., Summons, R. E. & Hinrichs, K.-U. Rapid resurgence of marine productivity after the Cretaceous–Paleogene mass extinction. *Science* **326**, 129–132 (2009).
8. Alegret, L., Thomas, E. & Lohmann, K. C. End-Cretaceous marine mass extinction not caused by productivity collapse. *Proc. Natl Acad. Sci. USA* **109**, 728–732 (2012).
9. Birch, H. S., Coxall, H. K., Pearson, P. N., Kroon, D. & Schmidt, D. N. Partial collapse of the marine carbon pump after the Cretaceous–Paleogene boundary. *Geology* **44**, 287–290 (2016).
10. Bown, P. Selective calcareous nannoplankton survivorship at the Cretaceous–Tertiary boundary. *Geology* **33**, 653–656 (2005).
11. Gallala, N., Zaghbib-Turki, D., Arenillas, I., Arz, J. A. & Molina, E. Catastrophic mass extinction and assemblage evolution in planktic foraminifera across the Cretaceous/Paleogene (K/Pg) boundary at Bidart (SW France). *Mar. Micropaleontol.* **72**, 196–209 (2009).
12. Hull, P. Life in the aftermath of mass extinctions. *Curr. Biol.* **25**, R941–R952 (2015).
13. Barnosky, A. D. et al. Merging paleobiology with conservation biology to guide the future of terrestrial ecosystems. *Science* **355**, eaah4787 (2017).
14. Hodgson, D., McDonald, J. L. & Hosken, D. J. What do you mean, 'resilient'? *Trends Ecol. Evol.* **30**, 503–506 (2015).
15. Varol, O. in *Nannofossils and their Applications. British Micropaleontological Society Series* (eds Crux, J. A. & van Heck, S. E.) 267–310 (Ellis Horwood, 1989).
16. Pospichal, J. J. in *The Cretaceous–Tertiary Event and Other Catastrophes in Earth History. Geological Society of America Special Paper 307* (eds Ryder, G. et al.) 335–360 (Geological Society of America, 1996).
17. Vellekoop, J. et al. Rapid short-term cooling following the Chicxulub impact at the Cretaceous–Paleogene boundary. *Proc. Natl Acad. Sci. USA* **111**, 7537–7541 (2014).
18. MacLeod, K. G., Quinton, P. C., Sepúlveda, J. & Negra, M. H. Postimpact earliest Paleogene warming shown by fish debris oxygen isotopes (El Kef, Tunisia). *Science* **360**, 1467–1469 (2018).
19. Sprain, C. J. et al. The eruptive tempo of Deccan volcanism in relation to the Cretaceous–Paleogene boundary. *Science* **363**, 866–870 (2019).
20. Zachos, J. C., McCarren, H., Murphy, B., Röhl, U. & Westerhold, T. Tempo and scale of late Paleocene and early Eocene carbon isotope cycles: implications for the origin of hyperthermals. *Earth Planet. Sci. Lett.* **299**, 242–249 (2010).
21. Gibbs, S. J. et al. Scaled biotic disruption during early Eocene global warming events. *Biogeosciences* **9**, 4679–4688 (2012).
22. Esmeray-Senlet, S. et al. Evidence for reduced export productivity following the Cretaceous/Paleogene mass extinction. *Paleoceanography* **30**, 718–738 (2015).
23. Alegret, L. & Thomas, E. Deep-sea environments across the Cretaceous/Paleogene boundary in the eastern South Atlantic Ocean (ODP leg 208, Walvis Ridge). *Mar. Micropaleontol.* **64**, 1–17 (2007).
24. Buesseler, K. O. & Boyd, P. W. Shedding light on processes that control particle export and flux attenuation in the twilight zone of the open ocean. *Limnol. Oceanogr.* **54**, 1210–1232 (2009).
25. Fischer, G. & Karakaş, G. Sinking rates and ballast composition of particles in the Atlantic Ocean: implications for the organic carbon fluxes to the deep ocean. *Biogeosciences* **6**, 85–102 (2009).
26. Hansen, B. et al. The size ratio between planktonic predators and their prey. *Limnol. Oceanogr.* **39**, 395–403 (1994).
27. Gerstel, J., Thunell, R. & Ehrlich, R. Danian faunal succession: Planktonic foraminiferal response to a changing marine environment. *Geology* **15**, 665–668 (1987).
28. Fuqua, L. M., Bralower, T. J., Arthur, M. A. & Patzkowsky, M. E. Evolution of calcareous nannoplankton and the recovery of marine food webs after the Cretaceous–Paleocene mass extinction. *Palaios* **23**, 185–194 (2008).
29. Barnosky, A. D. et al. Has the Earth's sixth mass extinction already arrived? *Nature* **471**, 51–57 (2011).
30. Cardinale, B. J. et al. Biodiversity loss and its impact on humanity. *Nature* **486**, 59–67 (2012).
31. Reich, P. B. et al. Impacts of biodiversity loss escalate through time as redundancy fades. *Science* **336**, 589–592 (2012).
32. Westerhold, T., Röhl, U., Donner, B. & Zachos, J. C. Global extent of early Eocene hyperthermal events: a new Pacific benthic foraminiferal isotope record from Shatsky Rise (ODP Site 1209). *Paleoceanogr. Paleoclimatol.* **33**, 626–642 (2018).

## METHODS

**Data reporting.** No statistical methods were used to predetermine sample size. The experiments were not randomized and investigators were not blinded to allocation during experiments and outcome assessment.

**Experimental design.** Our objective was to characterize the emergence of resilience in plankton communities in the aftermath of the K/Pg mass extinction and assess implications for higher trophic levels and biogeochemical cycling of the oceans. This palaeoecological study presents sustained very-high resolution sampling data (13 kyr) across a long duration (13 Myr), maintaining sampling intensity across both event and background stratigraphic intervals. To achieve this, we required a continuous, long time-series record from a single location that was representative of global patterns; all of these features were satisfied by the open-ocean gyre location of Ocean Drilling Program (ODP) Site 1209. The palaeogeography and overall oceanic setting varied little at this site across the 13-Myr-long record and calcareous nannoplankton provide the most-consistent, high-abundance fossil record of any plankton group. This site has an orbitally resolved age model, providing millennial sampling precision alongside high-resolution geochemical proxy records of palaeoenvironmental change[32]. Furthermore, the site is far from the Yucatan bolide impact and samples the dominant ocean basin of the early Cenozoic, providing the potential to track the recovery of a marine ecosystem on a quasi-globally representative basis (see 'cGENIE Earth system modelling and palaeo-hydrographic location of Site 1209').

**Sampling strategy.** The nannoplankton assemblage data come from an approximately 54-m section of the composite splice[33] at ODP Site 1209 Shatsky Rise, in the palaeo-subequatorial Pacific Ocean (Extended Data Fig. 1). We obtained 981 samples at approximately 13-kyr intervals, extending from the K/Pg boundary (66 Myr ago) to the Palaeocene–Eocene Thermal Maximum (PETM; 56 Myr ago), and overlap the Eocene record of a previously published study[21], giving a 13-Myr-long record in total. The ages assigned to each sample use age models constructed for ODP Site 1209 as described previously[34,35] (option 2) and as updated in a previous study[32]. The age model uses tie-points in the $\delta^{13}$C data that are correlated with the orbitally tuned stratigraphy of ODP Site 1262, as summarized previously[20,32].

**Assemblage data.** Smear slides were prepared for nannofossil observation using standard techniques[36]. Assemblage data (Extended Data Fig. 3) are based on statistically significant counts of 500 to 1,000 nannofossil liths per sample across a minimum of 10 fields of view, following identical count and taxonomic protocols of the previously published Eocene record[21] (~218.5 to 208.0 revised metres composite depth; around 56–53 Myr ago). The assemblage data were counted to generic level, with some additional division into useful morphogroups (for example, determined by genus and size, see 'Selected taxonomic notes'). Taxonomy generally follows previously published studies[37–39]. Visual assessment of preservation, as well as quantitative counts of lith fragmentation and presence of delicate lith structures, indicates that the nannofossils are generally well-preserved but show some signs of etching and minor overgrowth, as is typical for carbonate-rich deep-sea sediments. Etching of delicate central-area structures—particularly of Prinsiaceae specimens—is common, but does not inhibit identification to genus level. Although there is always some degree of variation in preservation quality across a long time series such as this, our observations show that dissolution has not disproportionately distorted the assemblage character at any particular level or interval. This includes the hyperthermals during which carbonate dissolution often increases[21] and, conversely, the immediate post-K/Pg interval for which indicators suggest enhanced preservation[40]. Of note, the later absence of the exclusively small taxa—*Neobiscutum*, small *Praeprinsius* and *Futyania*—is an evolutionary signal confirmed in sections worldwide[15], rather than a preservation artefact. There is evidence of reworking of Cretaceous taxa immediately above the K/Pg boundary at Site 1209 (Extended Data Fig. 3); however, these specimens have not been included in the relative abundance calculations.

**$\Sigma_{CV}$.** We used a range of approaches (see 'Dissimilarity and diversity metrics') to characterize community structure, but focus on the $\Sigma_{CV}$ time series as it best encapsulates the key trends in community variance and relationships with environmental perturbation. The $\Sigma_{CV}$ method is an analytical technique that is independent of taxonomic composition, and enables efficient collection and integration of large amounts of abundance data, giving equal weighting to each of the taxa included[21]. When applied to microfossil data, it highlights the nuances of biotic response across a very broad spectrum of perturbations. We follow the same procedure as published previously[21], but without using the SiZer smoothing step as we are not comparing datasets from different sources. First, the assemblage data, collected from samples taken every approximately 5.5 cm (equivalent to every 13 kyr) were placed on the age scale and linearly resampled using AnalySeries version 2.0[41] to provide consistent 13-kyr spacing between data points. Second, we determined which taxa would be included for subsequent $\Sigma_{CV}$ analysis. Because the 13-Myr-long record includes considerable evolution in community taxon composition, we divided the section into Myr-long bins and determined the most abundant and consistently present (>65% of samples) taxa in each. This resulted in the selection of 8 taxa across each

bin—a relatively low number because of the low diversity in the early Danian, but which represents >95% of the total population in each sample. We then followed the $\Sigma_{CV}$ method of calculating coefficients of variation summed across these taxa using a sliding window duration of 150 kyr. As the $\Sigma_{CV}$ metric quantifies the levels of variance across multiple taxa, our use of the term stability refers to consistent and low levels of change in the abundance distribution across the main taxa. The term stability is used in ecology in a myriad of ways; however, in this case we use a simple and intuitive definition of stability as meaning a system with low variability (that is, little deviation from its average state[42])—a definition that we think is most directly applicable to geological time-series data.

**$\Sigma_{CV}$ sensitivity tests.** We applied a range of sensitivity tests to the $\Sigma_{CV}$ metric record, examining the effects of sample window duration, taxon dominance, ancestry, fossil preservation, sedimentation rate and hiatuses[21]. We analyse the impact of varying window duration in Extended Data Fig. 5 and show how variance is packaged through time, as well as any differences that result from analysing the data in time versus age domains. $\Sigma_{CV}$ increases with increasing window duration in the lower Danian, indicating that the window is capturing additional variance that is spread throughout the interval. By contrast, $\Sigma_{CV}$ decreases with increasing window duration across the PETM and Eocene Thermal Maximum 2, indicating focused variance, with little additional variance in the broader time window diluting the signal. We explored the effect of shared ancestry because our analyses give equal weight to each taxonomic unit, potentially introducing artificially high variance. We tested for this by re-analysing the data using two additional models of shared ancestry developed from our genus-level stratophenetic tree (Extended Data Fig. 6). Sensitivity of the $\Sigma_{CV}$ metric decreases as more genera are grouped, damping levels of variance (Extended Data Fig. 7), particularly when merging abundant genera from the same family (the highly conservative ancestry, option 1). However, the main patterns of variance still remain as robust features (as they do in the dissimilarity index described in 'Dissimilarity and diversity metrics'), particularly when more branches of the tree are conserved (the moderately conservative ancestry, option 2).

**Dissimilarity and diversity metrics.** We calculated additional metrics of assemblage structure, namely Bray–Curtis dissimilarity (a metric that highlights structural differences in abundance and composition), the Simpson's index (an evenness/dominance metric that incorporates abundance distribution and taxic richness) and the standard deviation (variance) of the Simpson's index (Extended Data Fig. 2). Bray–Curtis dissimilarity was performed on the maximum and minimum abundance values across the 11 samples within each moving 150-kyr window, returning the maximum dissimilarity value. The values were plotted for each moving 150-kyr window across the time series. Bray–Curtis dissimilarity is sensitive to taxonomic turnover (shown by increasing values with increasing window size; Extended Data Fig. 5b), but the impact is minimized using the 150-kyr window, as species turnover is low. The standard deviation (variance) of the Simpson's index was calculated from the 11 samples in each 150-kyr moving window. The $\Sigma_{CV}$ and Bray–Curtis dissimilarity time-series patterns are very similar ($R^2$ of 63%); however, the Bray–Curtis dissimilarity record differs in the amplitude of variation across background intervals because it is influenced—to varying degrees throughout the time series—by rare taxa. Simpson's index is also highly sensitive to the abundances of rarer taxa and the rare, variable occurrences of taxa close to their appearance and/or disappearance.

**Species richness estimates.** The new species richness diversity data are an update of the global compilation of a previous study[3]. We added new taxa described since 2004, increased the temporal resolution to 250-kyr stratigraphic bins for the entire dataset (Fig. 1 and Extended Data Fig. 4) and 100-kyr bins for the Danian (Fig. 3), and present the data on the GTS2012 timescale[43]. The species richness is the total number of taxa that occur for some part of, or throughout, each stratigraphic bin. Species richness estimates are dependent on the bin duration, hence the difference between the values in Figs. 1f and 3d.

**Cell size and volume.** Estimates of average cell volume (Fig. 3 and Extended Data Fig. 4) are based on mean cell size per taxon weighted according to their abundance in the community at any given time:

$$\text{Average cell volume} = ((([(\%T1_{ra} \times \Theta T1_{av}) + (\%T2_{ra} \times \Theta T2_{av}) + (\%T3_{ra} \times \Theta T3_{av}) \dots . Tn]/2)^3)/3 \times 4\pi)/\Sigma\%T_{ra}$$

in which $\%T1_{ra}$ is the percentage relative cellular abundance of the taxon (T) in the total assemblage and $\Theta T1_{av}$ is the average cell diameter of the taxon.

Cell diameter uses the internal diameter of coccospheres and cellular abundances were estimated by dividing the relative abundance of liths that are present by the average number of liths per cell. Coccosphere size and lith number for each taxon use (i) direct coccosphere measurements from coeval samples from ODP Site 1209 (Shatsky Rise) and Integrated Ocean Drilling Program (IODP) sites 1403 and 1407 (North Atlantic), and from published scanning electron microscopy images of coccospheres[44,45]; (ii) coccolith measurements from these same samples

converted to estimated cell size (and associated lith number) based on taxon-specific relationships between lith size, lith number and cell size determined from Palaeogene taxa within the same genus or family[46]; or (iii) estimates using modern analogues[46] (details in Extended Data Table 1). For the calcareous dinocysts, we took a conservative estimate of cell diameter of 20 μm, based on light microscope and scanning electron microscopy images of complete dinocyst coverings from the lowermost Danian of ODP Site 1210 (Shatsky Rise) and divided raw calcareous dinocyst fragment counts by 12, as an estimate of how many fragments constitute a whole cell. Estimated PIC per cell uses the least-squares linear regression between cell volume and cell PIC of figure 4c of a previously published study[46].

**cGENIE Earth system modelling and palaeo-hydrographic location of Site 1209.** We illustrate the palaeo-hydrographic location of Site 1209 using the cGENIE Earth system model. In this simulation, cGENIE is configured with late Maastrictian boundary conditions of continental configuration, bathymetry and wind stress as described previously[47]. Additionally, the solar constant is reduced appropriate for the time (66 Myr ago) and atmospheric $CO_2$ is set to 1,112 ppm ($4\times$ pre-industrial levels). We take the 10-kyr spin-up described previously[47] and run this on for 10 more years, showing the results of the last year of the 10-year follow-on experiment in Extended Data Fig. 1 as an annual average. ODP Site 1209 was slightly to the north (around 8°) of the palaeo-Equator 66 Myr ago (Extended Data Fig. 1a), lying towards the edge of an ocean current field that is circumequatorial (Extended Data Fig. 1b) and links the major ocean basins. In the simulated late Maastrictian climate, temperatures are not more than about 6 °C cooler than those at the location of Site 1209 (35 °C), nor do they exceed this value, anywhere along the flow path by more than a few degrees Celsius. Furthermore, from simple visual inspection of the cGENIE simulations (Extended Data Fig. 1), the deflection of the circumequatorial current south of China and southeast Asia to latitudes of around 10° S and interaction with the South Pacific subtropical gyre suggest the potential for considerable surface-water mixing to occur between the hemispheres. We conclude from this that Site 1209 is likely to be sampling the same tropical and partly sub-tropical plankton communities that occur in all major ocean basins and both hemispheres. The area of connected waters in the 28–38 °C range is over 50% of the global ocean surface. The obvious exceptions to this global connectivity are the Arctic (being characterized by much cooler temperatures) and the South Atlantic (which exchanges with the Pacific primarily only to the south of Africa, with the cooler water regime in this ocean gateway representing a potential barrier to the mixing of tropical plankton communities globally).

**Palaeogene climate events.** A number of important climate events occur during the 13-Myr study interval, including named transient events marked by isotopic excursions and identified on Figs. 1 and 3, with further details provided in Extended Data Table 2. These are mainly global warming hyperthermal events identified by carbon and oxygen isotope excursions and associated deep-sea carbonate dissolution. Events I2, I1, H2, H1 and the PETM were recognized at ODP Site 1209 by examination of benthic and bulk carbon isotope values and magnetic susceptibility data, following previous studies[21,32]. The PETM was also identified in benthic carbon isotope values and X-ray fluorescence (XRF) Fe intensity data, as described previously[35,48]. The Palaeocene Carbon Isotope Maximum, Early Late Palaeocene Event, Latest Danian Event and the K/Pg boundary were identified in benthic carbon isotope values and XRF Fe intensity data[48], and the positions of the Palaeocene Carbon Isotope Maximum and Latest Danian Event were verified against records from ODP Site 1262[49]. The lower C29n and Dan-C2 events are not clear in the benthic carbon isotope data at ODP Site 1209[48], but were identified according to a previous publication[50], in which it was suggested that the peaks in magnetic susceptibility[51] and XRF Fe intensity[35] identified as Pa2 and Pa1 in the study[35] correlate with the lower C29n and Dan-C2 events, respectively. The position of Dan-C2 is consistent with estimates for the timing of this event[34,52].

**Relationships between variance and carbon isotope excursion magnitude.** For each climate event, we used the carbon isotope excursion (CIE) magnitude as a proxy for the level of environmental perturbation, as illustrated by the scaling of temperature change with CIE size for several of the Eocene hyperthermals[53]. For the purposes of comparing environmental perturbation and $\Sigma_{CV}$ (Fig. 2), we plotted the magnitude of CIE using a combination of size of excursion as recorded at ODP Site 1209 and the magnitude of excursion estimated from published bulk carbon isotope data at globally distributed sites (Extended Data Table 2). We used the maximum recorded excursion, except in cases in which this was inconsistent with other available data. In addition to values from bulk carbon isotope data (consistent with a previous study[21]), we took into account available benthic CIE values, which are arguably preferable for resolving global signals[54]. The value of carbon isotopes used for plotting non-event $\Sigma_{CV}$ data points in Fig. 2 uses the deviation of the carbon isotope value from the detrended running average (using an 11-point running average through non-event-only values), for a data point every approximately 150 kyr between climatic events. We regressed first differences in $\Sigma_{CV}$ and first differences in CIE magnitude (Fig. 2b) to statistically explore the relationship between community stability and climate change across this 13-Myr-long interval,

using a generalized least-squares framework (gls function in the nlme library of R) that applies best-fit models that incorporate heteroscedastic (non-constant variance with the mean) and temporally autocorrelated (time series) errors.

**Milestones.** *Biological pump reboot and recovery.* Carbon isotope records of benthic and planktonic foraminifera from Walvis Ridge (southern Atlantic), adjusted to account for vital effects and ecology, show a crash in surface- to deep-water carbon isotope gradients at the mass extinction level and indicate that transfer of organic matter to the deep sea by the biological pump was severely perturbed[9]. These records show that vertical gradients were close to zero for the initial 0.3 Myr after the extinction, after which they slowly increased to attain pre-extinction levels at around 1.77 Myr after the extinction. This is interpreted as evidence that the duration of weakened biological pumping was no longer than 1.77 Myr long[9], providing an estimate for full biological pump recovery.

*Photosymbiosis and depth partitioning in planktonic foraminifera.* On the basis of reconstructions of the palaeoecology of planktonic foraminifera using the oxygen and carbon stable isotopes of their shells, the appearance of photosymbiosis and expansion of depth partitioning both occur around 2.5 Myr after the mass extinction[9,55].

*Appearance of oligotrophic coccolithophores.* The first appearance of early fasciculiths and sphenoliths represent the earliest representatives of the Discoasterales group, which is largely characterized by oligotrophic taxa[28]. The earliest representatives, *Fasciculithus magnus* and *F. magnicordis*, appear around 63 Myr ago (as described here and previously[56]) with other fasciculiths and sphenoliths following soon after (62.13 and 61.98 Myr ago, respectively, according to a previously published study[43]).

**Selected taxonomic notes.** *Praeprinsius* as described here includes very small (<3 μm) circular to subcircular specimens of *Praeprinsius tenuiculum*. *Praeprinsius* is considered a synonym of *Prinsius* by some, but we consider these groups to be morphologically distinct. For *Fasciculithus*, we use 'early fasciculiths' to include specimens that some[56–58] may now identify as *Gomphiolithus*, *Diantholitha* and *Lithoptychius*, while our main '*Fasciculithus*' group includes taxa that have been consistently classified within this genus, for example, *F. involutus* and *F. tympaniformis*. For sphenoliths, the earliest specimens of the genus *Sphenolithus*[56,59] are highly variable and we distinguish between the earliest incoming specimens (termed 'early sphenoliths') and the main generic group '*Sphenolithus*', which includes *S. primus/moriformis* and *S. anarrhopus*.

**Reporting summary.** Further information on research design is available in the Nature Research Reporting Summary linked to this article.

## Data availability

The datasets generated or analysed during this study are included as Source Data for Figs. 1–3.

33. Westerhold, T. & Röhl, U. Data report: revised composite depth records for Shatsky Rise Sites 1209, 1210, and 1211. In *Proc. Ocean Drilling Program, Scientific Results 198* (eds. Bralower, T. J. et al.) 1–26 (2006).
34. Dinarès-Turell, J., Westerhold, T., Pujalte, V., Röhl, U. & Kroon, D. Astronomical calibration of the Danian stage (Early Paleocene) revisited: settling chronologies of sedimentary records across the Atlantic and Pacific Oceans. *Earth Planet. Sci. Lett.* **405**, 119–131 (2014).
35. Westerhold, T. et al. Astronomical calibration of the Paleocene time. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **257**, 377–403 (2008).
36. Bown, P. R. & Young, J. R. in *Calcareous Nannofossil Biostratigraphy* (ed. Bown, P. R.) 16–28 (Chapman & Hall, 1998).
37. Perch-Nielsen, K. in *Plankton Stratigraphy* (eds. Bolli, H. M. et al.) 427–555 (Cambridge Univ. Press, 1985).
38. Bown, P. R. Paleocene calcareous nannofossils from Tanzania (TDP sites 19, 27 and 38). *J. Nannoplankton Res.* **36**, 1–32 (2016).
39. Young, J. R., Bown, P. R. & Lees, J. A. Nannotax 3. http://www.mikrotax.org/Nannotax3/; accessed 2015–2019.
40. Bown, P. R. et al. A Paleogene calcareous microfossil Konservat-Lagerstätte from the Kilwa Group of coastal Tanzania. *Geol. Soc. Am. Bull.* **120**, 3–12 (2008).
41. Paillard, D., Labeyrie, L. & Yiou, P. Macintosh program performs time-series analysis. *Eos* **77**, 379 (1996).
42. Cleland, E. E. Biodiversity and ecosystem stability. *Nature Education Knowledge* **3**, 14 (2011).
43. Gradstein, F. M., Ogg, J. G., Schmitz, M. D. & Ogg, G. M. *The Geologic Time Scale 2012* 1st edn (Elsevier, 2012).
44. Mai, H., Perch-Nielsen, K., Willems, H. & Romein, A. J. T. Fossil coccospheres from the K/T boundary section from Geulhemmerberg, The Netherlands. *Micropaleontology* **43**, 281–302 (1997).
45. Mai, H., Hildebrand-Habel, T., Perch-Nielsen, K. V. S. & Willems, H. Paleocene coccospheres from DSDP leg 39, Site 356, Sao Paulo Plateau, S Atlantic Ocean. *J. Nannoplankton Res.* **20**, 21–29 (1998).
46. Gibbs, S. J., Sheward, R. M., Bown, P. R., Poulton, A. J. & Alvarez, S. A. Warm plankton soup and red herrings: calcareous nannoplankton cellular communities and the Palaeocene–Eocene thermal maximum. *Phil. Trans. R. Soc. A* **376**, 20170075 (2018).

47. Schmidt, A. et al. Selective environmental stress from sulphur emitted by continental flood basalt eruptions. *Nat. Geosci.* **9**, 77–82 (2016).
48. Westerhold, T., Röhl, U., Donner, B., McCarren, H. K. & Zachos, J. C. A complete high-resolution Paleocene benthic stable isotope record for the central Pacific (ODP Site 1209). *Paleoceanography* **25**, PA2216 (2011).
49. Littler, K., Röhl, U., Westerhold, T. & Zachos, J. C. A high-resolution benthic stable-isotope record for the South Atlantic: implications for orbital-scale changes in Late Paleocene–Early Eocene climate and carbon cycling. *Earth Planet. Sci. Lett.* **401**, 18–30 (2014).
50. Coccioni, R. et al. The Dan-C2 hyperthermal event at Gubbio (Italy): global implications, environmental effects, and cause(s). *Earth Planet. Sci. Lett.* **297**, 298–305 (2010).
51. Shipboard Scientific Party. Site 1209. In *Proc. Ocean Drilling Program, Initial Reports 198* (eds Bralower, T. J. et al.) 1–102 (2002).
52. Westerhold, T., Röhl, U., Frederichs, T., Bohaty, S. M. & Zachos, J. C. Astronomical calibration of the geological timescale: closing the middle Eocene gap. *Clim. Past* **11**, 1181–1195 (2015).
53. Stap, L. et al. High-resolution deep-sea carbon and oxygen isotope records of Eocene thermal maximum 2 and H2. *Geology* **38**, 607–610 (2010).
54. Zeebe, R. E., Westerhold, T., Littler, K. & Zachos, J. C. Orbital forcing of the Paleocene and Eocene carbon cycle. *Paleoceanography* **32**, 440–465 (2017).
55. Birch, H., Coxall, H. K. & Pearson, P. N. Evolutionary ecology of Early Paleocene planktonic foraminifera: size, depth habitat and symbiosis. *Paleobiology* **38**, 374–390 (2012).
56. Monechi, S., Reale, V., Bernaola, G. & Balestra, B. The Danian/Selandian boundary at Site 1262 (South Atlantic) and in the Tethyan region: biomagnetostratigraphy, evolutionary trends in fasciculiths and environmental effects of the latest Danian event. *Mar. Micropaleontol.* **98**, 28–40 (2013).
57. Aubry, M.-P., Bord, D. & Rodriguez, O. New taxa of the order Discoasterales Hay 1977. *Micropaleontology* **57**, 269–288 (2011).
58. Monechi, S., Reale, V., Bernaola, G. & Balestra, B. Taxonomic review of early Paleocene fasciculiths. *Micropaleontology* **58**, 351–365 (2012).
59. Agnini, C. et al. High-resolution nannofossil biochronology of middle Paleocene to early Eocene at ODP Site 1262: implications for calcareous nannoplankton evolution. *Mar. Micropaleontol.* **64**, 215–248 (2007).
60. Cramer, B. S., Wright, J. D., Kent, D. V. & Aubry, M.-P. Orbital climate forcing of $\delta^{13}C$ excursions in the late Paleocene–early Eocene (chrons C24n–C25n). *Paleoceanography* **18**, 1097 (2003).
61. Nicolo, M. J., Dickens, G. R., Hollis, C. J. & Zachos, J. C. Multiple early Eocene hyperthermals: their sedimentary expression on the New Zealand continental margin and in the deep sea. *Geology* **35**, 699–702 (2007).
62. Lourens, L. J. et al. Astronomical pacing of late Palaeocene to early Eocene global warming events. *Nature* **435**, 1083–1087 (2005).
63. Zachos, J. C. et al. A transient rise in tropical sea surface temperature during the Paleocene–Eocene thermal maximum. *Science* **302**, 1551–1554 (2003).
64. Zachos, J. C. et al. Rapid acidification of the ocean during the Paleocene-Eocene thermal maximum. *Science* **308**, 1611–1615 (2005).
65. Bernaola, G. et al. Evidence of an abrupt environmental disruption during the mid-Paleocene biotic event (Zumaia section, western Pyrenees). *Bull. Geol. Soc. Am.* **119**, 785–795 (2007).
66. Quillévéré, F., Aubry, M.-P., Norris, R. D. & Berggren, W. A. Paleocene oceanography of the eastern subtropical Indian Ocean: an integrated magnetobiostratigraphic and stable isotope study of ODP hole 761B (Wombat Plateau). *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **184**, 371–405 (2002).
67. Arenillas, I., Molina, E., Ortiz, S. & Schmitz, B. Foraminiferal and $\delta13C$ isotopic event-stratigraphy across the Danian–Selandian transition at Zumaya (northern Spain): chronostratigraphic implications. *Terra Nova* **20**, 38–44 (2008).
68. Bornemann, A. et al. Latest Danian carbon isotope anomaly and associated environmental change in the southern Tethys (Nile Basin, Egypt). *J. Geol. Soc. Lond.* **166**, 1135–1142 (2009).
69. Jehle, S., Bornemann, A., Deprez, A. & Speijer, R. Planktic foraminiferal response to the latest Danian event in the Pacific Ocean (ODP Site 1210). *Geophysical Research Abstracts* **17**, EGU2015–9115 (2015).
70. Schmitz, B. et al. High-resolution $\delta^{13}C$, $\delta^{18}O$, foraminifera and nannofossil profiles across the latest Paleocene benthic extinction event at Zumaya, Spain. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **133**, 49–68 (1997).
71. Schmitz, B., Molina, E. & von Salis, K. The Zumaya section in Spain: a possible global stratotype section for the Selandian and Thanetian stages. *Newsl. Stratigr.* **36**, 35–42 (1998).
72. Kroon, D. & Zachos, J. C. & Leg 208 Scientific Party. Leg 208 synthesis: Cenozoic climate cycles and excursions. In *Proc. Ocean Drilling Program, Scientific Results 208* (eds Kroon, D. et al.) 1–55 (2007).
73. Quillévéré, F., Norris, R. D., Kroon, D. & Wilson, P. A. Transient ocean warming and shifts in carbon reservoirs during the early Danian. *Earth Planet. Sci. Lett.* **265**, 600–615 (2008).
74. van Heck, S. E. & Prins, B. A refined nannoplankton zonation for the Danian of the central North Sea. *Abh. Geol. Bundesanstalt* **39**, 285–303 (1987).
75. Sáez, A. G. et al. in *Coccolithophores – from Molecular Processes to Global Impact* (eds Thierstein, H. R. & Young, J. R.) 251–270 (Springer 2004).
76. Young, J. R., Geisen, M. & Probert, I. A review of selected aspects of coccolithophore biology with implications for palaeobiodiversity estimation. *Micropaleontology* **51**, 267–288 (2005).
77. Young, J. R., Liu, H., Probert, I., Aris-Brosou, H. & de Vargas, C. Morphospecies versus phylospecies concepts for evaluating phytoplankton diversity: the case of the coccolithophores. Cryptogam. *Algol.* **35**, 353–377 (2014).

**Extended Data Fig. 1 | Location of ODP Site 1209 (black star) with respect to model-simulated late Cretaceous major ocean current and circulation patterns. a**, Barotropic stream function (Sv) simulated in a late Cretaceous configuration of the cGENIE Earth system model[47]. **b**, Surface ocean current field (black arrows) for the same circulation state as **a** overlaid on annual average sea surface temperature (SST) (colours). Scale for current vectors on the right, along with a truncated temperature scale to highlight the distribution of comparable temperature regimes. Red arrows illustrate inferred flow paths relevant to the position of ODP Site 1209 (marked by a star).
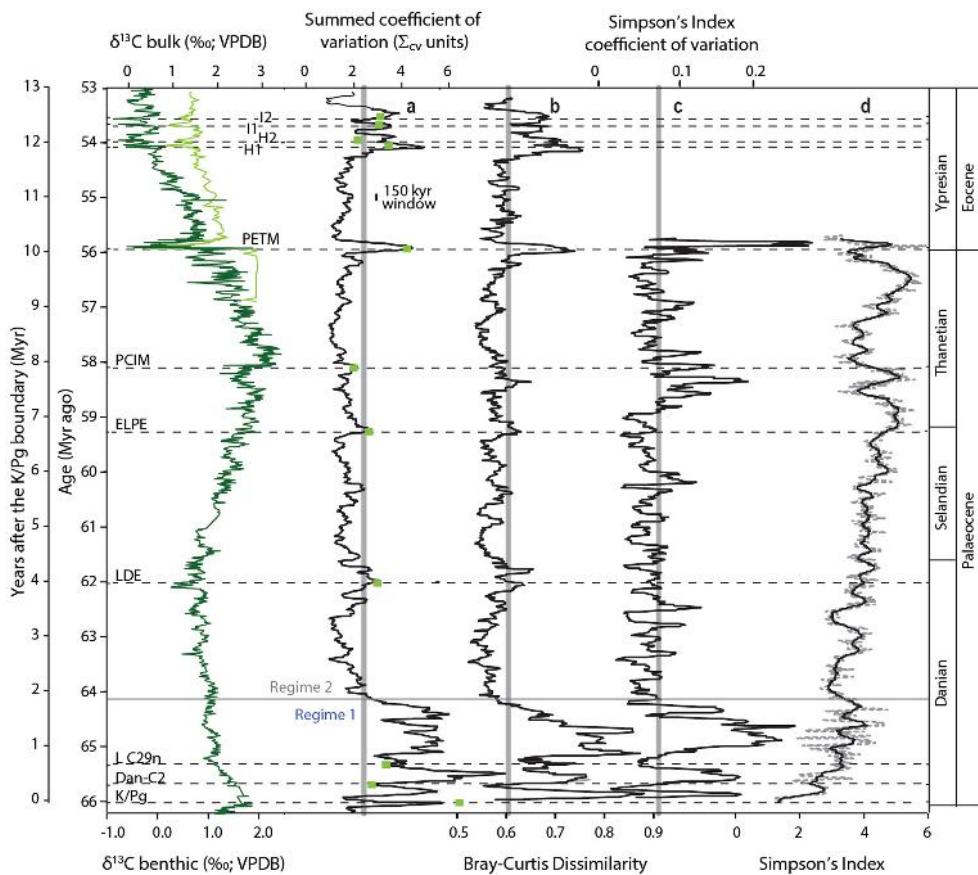
**Extended Data Fig. 2 | Comparison of community structure metrics.**
**a–d**, Left, δ13C bulk and stable isotopes as in Fig. 1a. Dark green, benthic[32]; light green, bulk. Downcore plots of $\Sigma_{CV}$ (**a**), Bray–Curtis dissimilarity (**b**), Simpson's index (**d**; grey dashed lines; black line indicates the 150-kyr moving average) and the variance (in 150-kyr windows) in the Simpson's index (**c**). Vertical grey lines in **a** and **b** show the level of background inferred from rank order plots of these data. All four metrics ($\Sigma_{CV}$, Bray–Curtis dissimilarity, Simpson's index and variance in the Simpson's index) show a distinction of volatility between early Danian regime 1 ($n = 137$ data points) and regime 2 (the rest of the record, $n = 861$ data points). For example, the Wilcoxon rank-sum ($W$) value for the Simpson's record was $W = 46,646$; $P < 0.001$ on first differences with 95% confidence limits of $-0.013$, $-0.006$. A $W$ value of zero would support a null hypothesis. The test was two-sided. The Simpson's index shows a diversity minimum

in the earliest Danian and then a rapid increase and steady long-term trend towards more diverse, more even communities, but with high variability in the early Danian. This fluctuation in the Simpson's index, as recorded by the variance of the record (**c**) shows similar patterns to Bray–Curtis dissimilarity and $\Sigma_{CV}$ with high variance in the early Danian before dropping down. The variance in the Simpson's index also shows high background fluctuations and a sustained increase in amplitude of fluctuations around the isotope shift in the Palaeocene Carbon Isotope Maximum, reflecting oligotroph diversification, which the Simpson's index shows strongly due to its higher sensitivity to rare taxa. In effect, metric sensitivity to the richness in taxa and rare taxa increased from **a** to **c** (from abundance variance to diversity variance). Note, the Simpson's index can only be calculated on full assemblage data and therefore the record extends only from 66 to 55.5 Myr ago.
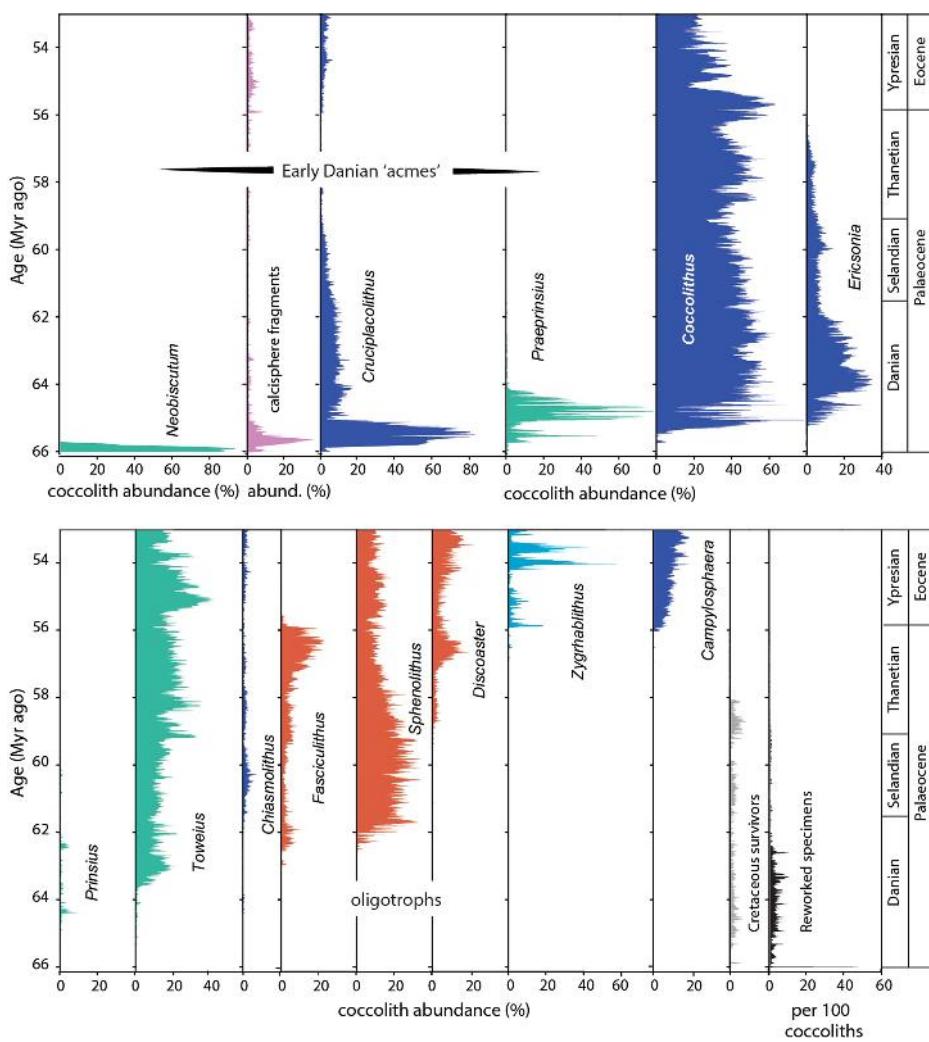
**Extended Data Fig. 3 | Relative abundance of key nannoplankton groups and abundance of reworked specimens per 100 nannofossils.** Relative abundance of coccoliths from all groups included in the $\Sigma_{CV}$ metric are shown, coloured according to clade (as in Fig. 1b) and ordered by stratigraphic appearance. Cretaceous survivor taxa were counted as individual species but have been grouped together here, comprising mostly *Zeugrhabdotus* with lower abundances of *Cyclagelosphaera*, *Markalius* and *Neocrepidolithus*.
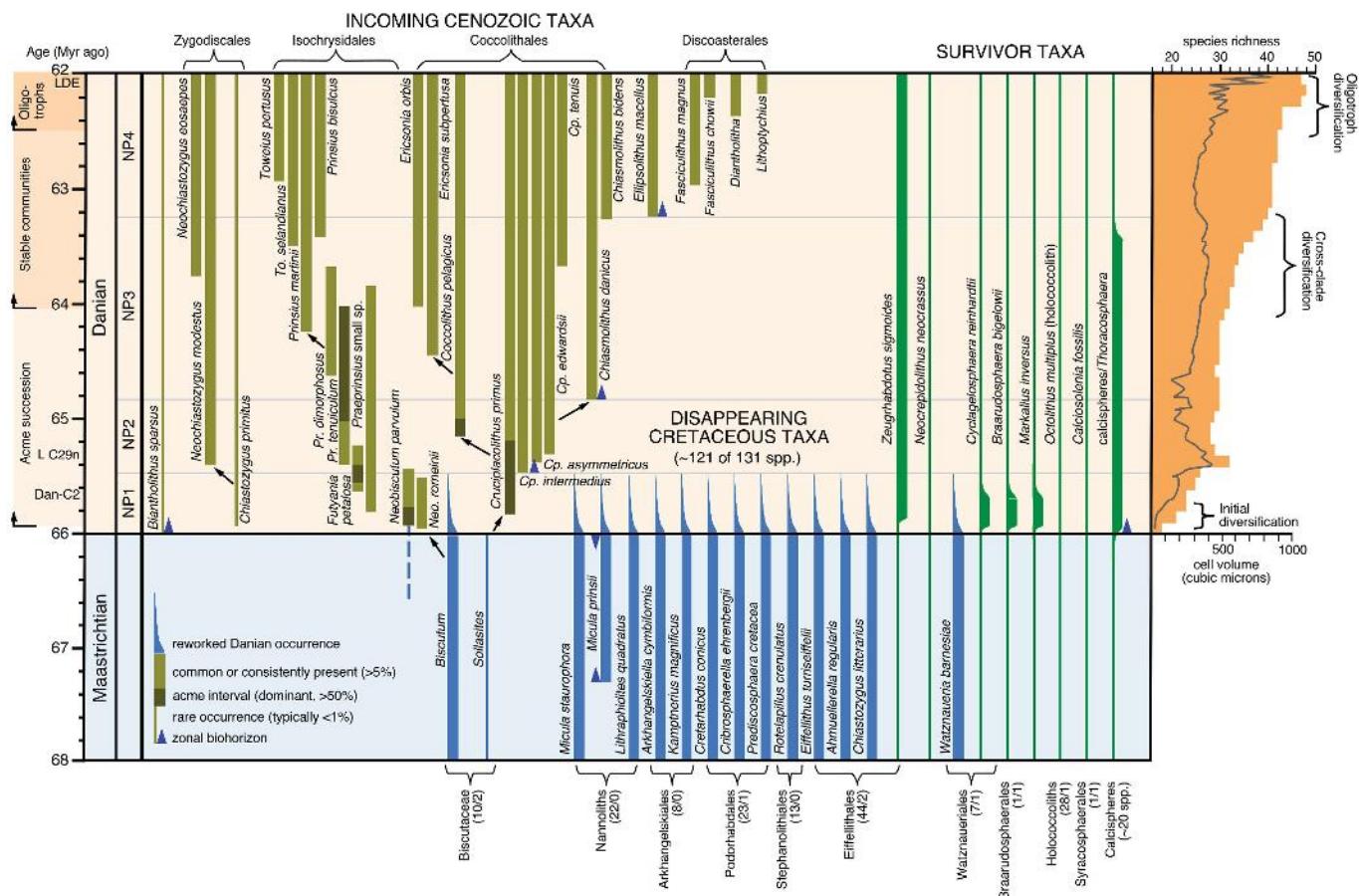
**Extended Data Fig. 4 | Calcareous nannoplankton across the K/Pg boundary.** Stratigraphic distribution of important species grouped as incoming (brown), survivor (green) or disappearing (blue) taxa. A subset of Cretaceous taxa is shown, with the latest Maastrichtian diversity for families shown alongside the number of survivors. Gradualistic evolutionary transitions indicated by close spacing and arrows indicate genus-level transitions. The nannoplankton data are primarily from our work but are largely consistent with published sources[10,15,16,37,74]. Diversity and cell-volume records are also shown in Fig. 3. *Cp.*, *Cruciplacolithus*; Dan-C2, L C29n, hyperthermals; LDE, Late Danian Event; *Neobisc.*, *Neobiscutum*; NP = nannofossil biozone; *Prae.*, *Praeprinsius*.
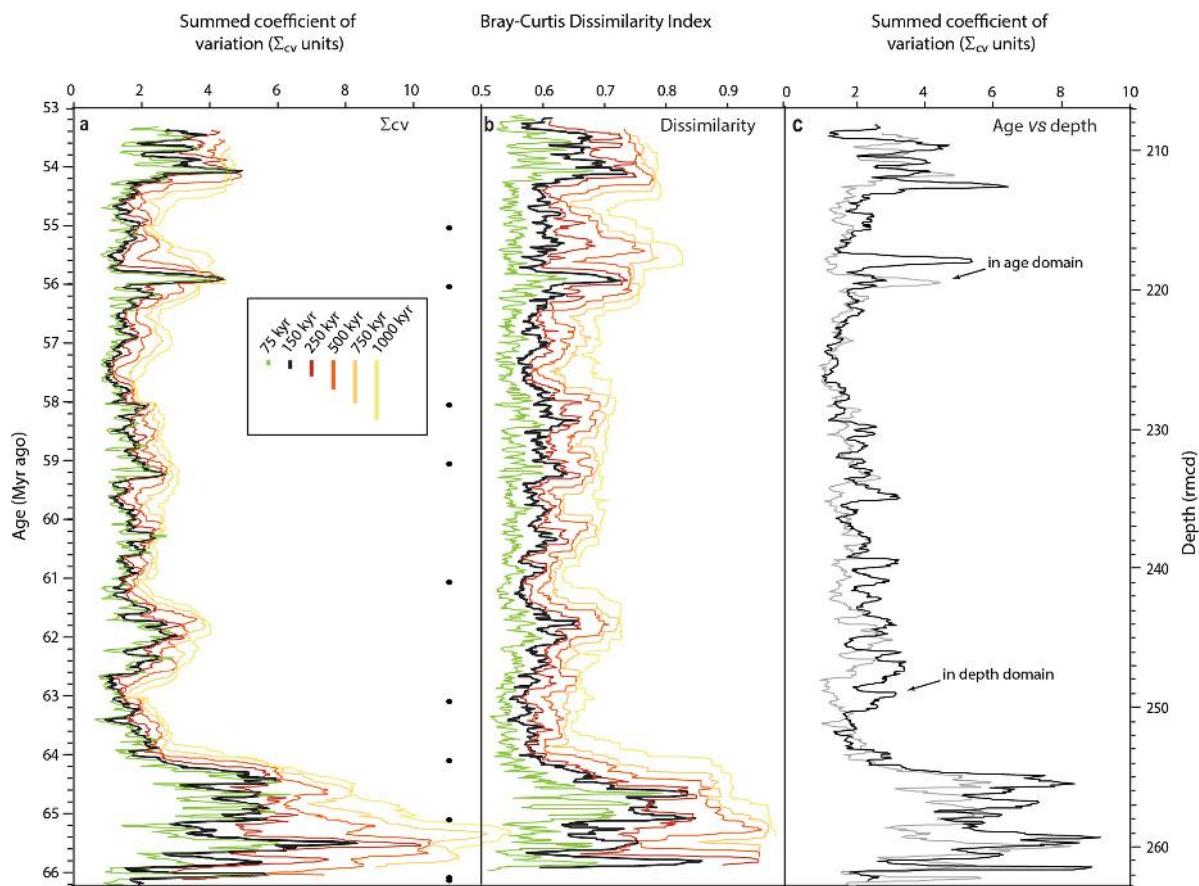
**Extended Data Fig. 5 | Effects of window duration and depth or age sampling on $\Sigma_{CV}$ and Bray–Curtis dissimilarity. a–c,** Influence of different window duration (75–1,000 kyr) (**a**) and sampling in either the depth or age domain (**c**) on $\Sigma_{CV}$, and effect of window duration on Bray–Curtis dissimilarity (**b**). **c,** Side-by-side results of $\Sigma_{CV}$ calculated using evenly spaced samples in either the depth domain or the age domain using a depth window duration of 60 cm, which is broadly equivalent to the 150-kyr time window. When $\Sigma_{CV}$ is plotted in the depth domain the main patterns are retained, indicating that no significant artefacts arise from the applied age model. The boundaries between the Myr sections, at which the taxa included in the $\Sigma_{CV}$ change (black dots), are marked on **a**. There are no obvious artefacts across the Myr windows with changes in the most abundant taxa.
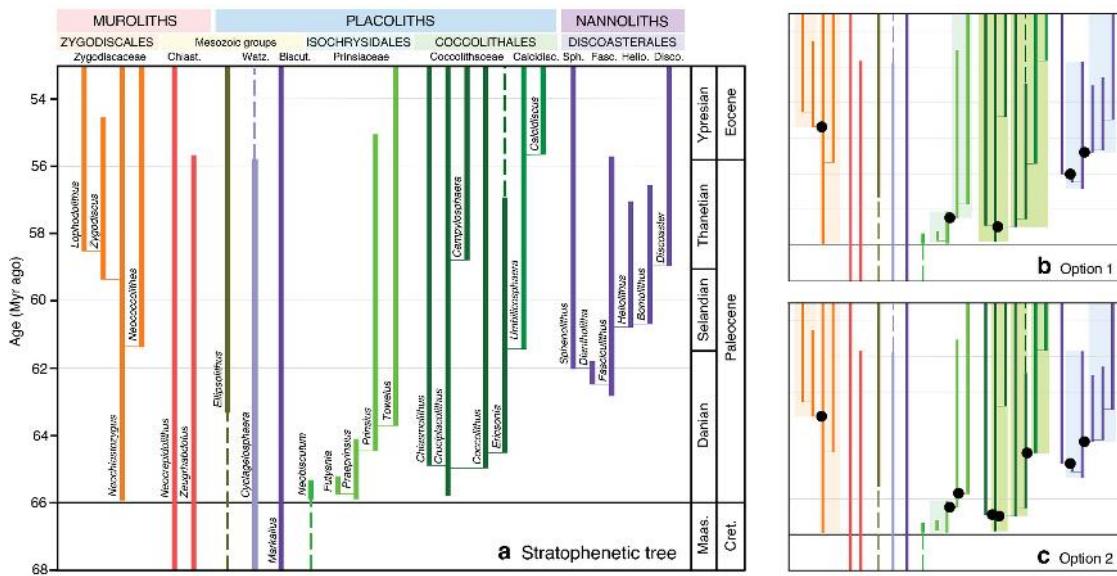
**Extended Data Fig. 6 | Phylogenetic models for the dominant Palaeocene nannoplankton. a–c**, Models range from a standard genus-level stratophenetic tree (**a**) through two successively conservative scenarios (**b**, **c**) grouping closely related taxa—that is, recently diverged taxa based on morphological and stratigraphic range data. Nannoplankton taxonomy is primarily based on the morphology and crystallographic ultrastructure of exoskeletal coccoliths but the addition of genetic data for modern taxa confirmed that this approach is robust[75–77]. Evolutionary models are stratophenetic, because we have high-quality stratigraphic information but lack the range of meaningful homologous morphological characters to allow a cladistic analysis. **a**, Genus-level phylogeny based on an extensive species-level stratophenetic tree. **b**, **c**, Different ancestry options used to test for artefacts and sensitivity in variance/dissimilarity that may result from equal weighting of closely related versus more-distantly related taxa. **b**, Ancestry model option 1 is highly conservative and merges major sub-family groups (shown by shaded boxes) around five nodes shown by black circles. **c**, Ancestry model option 2 merges the most-closely related genera (shaded boxes) around eight nodes.
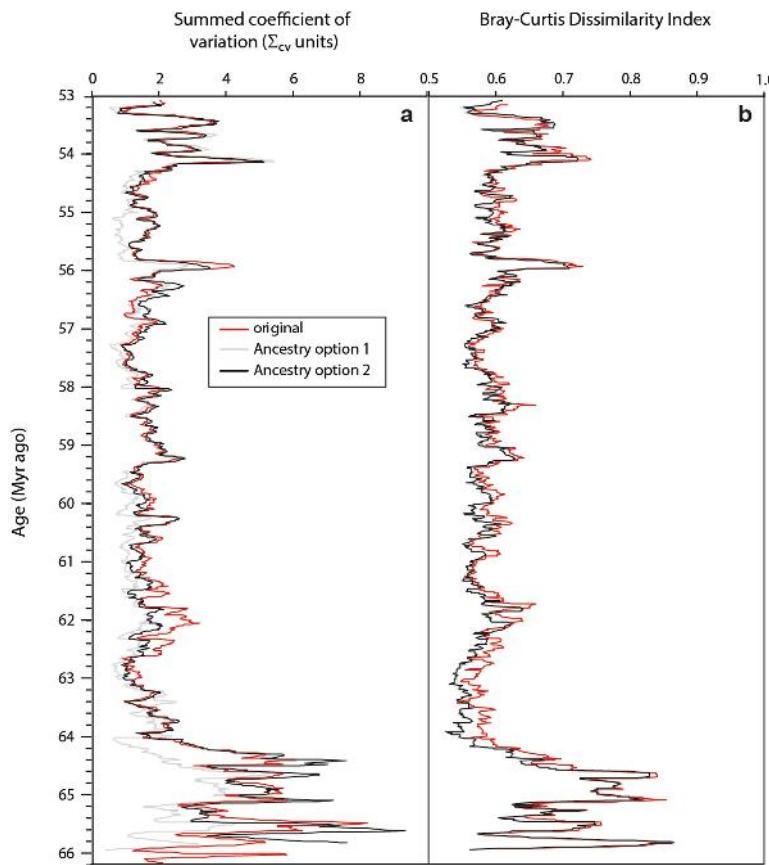
**Extended Data Fig. 7 | Influences of ancestry on $\Sigma_{CV}$ and Bray–Curtis dissimilarity. a**, **b**, Analyses of the influences of two additional models of shared ancestry (Extended Data Fig. 6a) on the $\Sigma_{CV}$ (**a**) and Bray–Curtis dissimilarity (**b**) datasets. In red, the original analysis in which each genus is weighted equally. In grey, analysis of the conservative ancestry model that merges genera into major sub-family groups (ancestry model option 1; Extended Data Fig. 6b). In black, analysis of the moderately conservative ancestry model (option 2; Extended Data Fig. 6c), which merges the most-closely related genera. The Bray–Curtis dissimilarity analysis shows very little sensitivity to variation in the taxonomic hierarchies. The $\Sigma_{CV}$ displays some sensitivity, particularly at the Late Danian Event (around 62 Myr ago); however, the main patterns are retained between the original and option 2. Some variance is lost in the less-realistic analysis of option 1, in which grouping of key genera that are found in the same families dampens the variance, in particular, in the early Danian. However, the values of early Danian variance still remain anomalously high compared to the rest of the record.

**Extended Data Table 1 | Summary of main biometric lith and cell parameters measured and reconstructed**

| Clade | Taxon | Lith number ($C_N$) | Cell size ($\Theta$, µm) | Cell volume (µm³) | Source |
|---|---|---|---|---|---|
| Cretaceous survivors | *Neocrepidolithus* | 32.5 | 16.5 | 2352 | Ref. 46 |
| | *Zeugrhabdotus* | 32.5 | 16.5 | 2352 | Ref. 46 |
| | *Markalius* | 12 | 8.3 | 297 | Pg coccospheres |
| | *Cyclagelosphaera* | 12 | 7.9 | 259 | $C_N$ from published SEM images; $\Theta$ estimated from lith measurements at Sites 1209, 1403 and published coccosphere images. Geometry consistent with Pg coccospheres |
| Coccolithales | *Chiasmolithus* | 9 | 11.0 | 698 | Pg coccospheres |
| | *Coccolithus* | 13 | 7.0-9.7 | 180-478 | $C_N$ from Pg coccospheres; $\Theta$ change through Danian estimated from lith measurements at Sites 1209 and 1403 using geometric relationship from Pg coccospheres |
| | *Cruciplacolithus* (small) | 20 | 4.2-8.0 | 39-268 | $C_N$ and $\Theta$ from Pg coccospheres and published coccosphere images; $\Theta$ change through Danian estimated from lith measurements from Sites 1209 and 1403 using geometric relationship from Pg coccospheres. |
| | *Cruciplacolithus* (large) | 13 | 6.6-11.5 | 151-796 | As for *Cruciplacolithus* (small), but considered to be more like *Chiasmolithus* |
| | *Ericsonia* | 13 | 7.6 | 226 | As for *Coccolithus* |
| Isochrysidales | *Neobiscutum* | 9 | 3.0 | 14 | Direct measurements of Danian coccospheres and estimates of $\Theta$ based on lith measurements from Sites 1209, 1403 and 1406. |
| | *Praeprinsius* | 18 | 3.8-5.2 | 29-74 | Direct measurements of Danian coccospheres and estimates of $\Theta$ change through Danian based on lith measurements from Sites 1209, 1403 and 1406. |
| | *Futyania* | 50 | 8.5 | 322 | Pg coccospheres and published coccosphere images |
| | *Prinsius* | 20 | 4.0-5.5 | 34-87 | Direct measurements of Danian coccospheres and estimates of $\Theta$ change through Danian based on lith measurements from Sites 1209, 1403 and 1406. |
| | *Toweius* | 7 | 4.7 | 31 | $C_N$ from Pg coccospheres; $\Theta$ estimated from lith measurements from Sites 1209 and 1403 using geometric relationship from Pg coccospheres. |
| Zygodiscales | *Neochiastozygus* | 32.5 | 16.5 | 2352 | Ref. 46 |
| Discoasterales | *Fasciculithus, Sphenolithus* | 42 | 21.7 | 5350 | Ref. 46 |
| Incertae | *Biantholithus* | 13 | 13.7 | 1337 | Pg coccospheres |
| Non-nannofossil | Calcisphere fragments | 16 | 20.0 | 4189 | Pg coccospheres |

$C_N$ is the number of coccoliths per cell, $\Theta$ is the cell diameter and $C_L$ is lith length. Sites referred to include ODP Site 1209 and IODP sites 1403 and 1407. 'Pg coccospheres' refers to new coccosphere measurements for the Palaeogene. Scanning electron microscopy images of published coccospheres are all from previously published studies[44,45].

**Extended Data Table 2 | Carbon isotope excursion events**

| Event | Depth (rmcd, splice) | Age (Myrs ago) | Size of CIE (‰) | References |
|---|---|---|---|---|
| I2 | 210.02 | 53.55 | **0.48** [0.1 − 0.6] | 21,60,61. |
| I1 | 210.60 | 53.67 | **0.65** [0.5 − 0.7] | 21,60,61. |
| H2 | 211.83 | 53.95 | **0.49** [0.2 − 0.6] | 21,53,60,61. |
| H1 | 212.48 | 54.05 | **1.5** [0.6 − 1.6] | 21,53,60−62. |
| PETM | 218.00 | 55.93 | **3.0** [2.4 − 3] | 63,64. |
| PCIM | 229.94 | 58.10 | 1.0 | 48,49. |
| ELPE | 235.00 | 59.27 | 0.75 | 48,49,65. |
| LDE | 247.69 | 62.03 | 1.0 | 48,66−71. |
| L C29n | 258.83 | 65.34 | 0.7 | 50,72. |
| Dan C2 | 260.11 | 65.71 | 1.3 | 50,72,73. |

Columns provide event nomenclature, depth in core at Site 1209, age and estimated size of the CIE. Values in parentheses show the range of CIEs from the literature[21,48–50,53,60–73] and the value in bold is the size of CIE used in Fig. 2. Event nomenclature follows references given in the Methods, depths (rmcd, revised metres composite depth) use the revised depth splice published previously[33] and the ages use the age model from a previous study[32].

# nature research

| | |
|---|---|
| Corresponding author(s): | Sarah Alvarez & Samantha Gibbs |
| Last updated by author(s): | Aug 13, 2019 |

# Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

## Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

| n/a | Confirmed | |
|---|---|---|
| ☐ | ☒ | The exact sample size ($n$) for each experimental group/condition, given as a discrete number and unit of measurement |
| ☐ | ☒ | A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly |
| ☐ | ☒ | The statistical test(s) used AND whether they are one- or two-sided<br>*Only common tests should be described solely by name; describe more complex techniques in the Methods section.* |
| ☐ | ☒ | A description of all covariates tested |
| ☐ | ☒ | A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons |
| ☐ | ☒ | A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals) |
| ☐ | ☒ | For null hypothesis testing, the test statistic (e.g. $F$, $t$, $r$) with confidence intervals, effect sizes, degrees of freedom and $P$ value noted<br>*Give P values as exact values whenever suitable.* |
| ☒ | ☐ | For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings |
| ☒ | ☐ | For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes |
| ☐ | ☒ | Estimates of effect sizes (e.g. Cohen's $d$, Pearson's $r$), indicating how they were calculated |

*Our web collection on statistics for biologists contains articles on many of the points above.*

## Software and code

Policy information about availability of computer code

| | |
|---|---|
| Data collection | *Provide a description of all commercial, open source and custom code used to collect the data in this study, specifying the version used OR state that no software was used.* |
| Data analysis | AnalySeries version 2.0, R software environment |

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

## Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Raw data associated with the main text figures is submitted (one per figure, Figures 1-3) with this submission. Much of this data was originally deposited with Pangaea (data submission 2018-09-04T08:40:33Z) when the manuscript was first submitted. However, now all the datasets are going to linked directly to the paper and so there is no need for the Pangaea archiving.

# Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☐ Life sciences     ☐ Behavioural & social sciences     ☒ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

# Ecological, evolutionary & environmental sciences study design

All studies must disclose on these points even when the disclosure is negative.

| | |
|---|---|
| Study description | The study targeted the interval after the mass extinction at the Cretaceous-Paleogene boundary that decimated marine and terrestrial biota. The aim was to track, at high resolution, recovery not just in the immediate aftermath of the extinctions but as the ocean ecosystem regained stability and resilience, and rebuilt diversity across the next ~13 million years through to the early Eocene - bridging gaps between previous studies and estimates/definitions of 'recovery'. Calcareous nannoplankton give the opportunity to acquire near-continuous assemblage data across this interval, with consistently high specimen numbers and good preservation, for a biogeochemically relevant group of organisms that provide us with insights into the health of the base of the ocean food web. ODP Site 1209 (Shatsky Rise, central Pacific) was chosen because we wanted to generate a long time-series record from a single site with no breaks in stratigraphy and which had a generally similar palaeogeography and oceanic setting throughout this interval. Also, the Pacific Ocean was the largest oceanic ecosystem during this time interval and ODP Site 1209 gives us close to a globally representative plankton picture as its position relative to surface circulation patterns at this time means it has communication with the other major ocean basins. |
| Research sample | The central dataset utilised in this study comprised a total of 990 samples at approximately 13 kyr spacing across 54 metres of composite core splice, corresponding to 13 million years. The samples are subsamples (toothpick samples, i.e., small scrapings) from deep sea cores drilled by the International Ocean Discovery Program. The samples are carbonate rich mudrocks from the central Pacific and are comprised predominantly of calcareous nannoplankton fossils. Assemblage data from the youngest 238 samples from the latest Paleocene through to Eocene are published in Gibbs et al. (2012) and the data from the oldest 11 samples from the Cretaceous are published in Bown (2005). The assemblage data from the intervening 741 samples are new to this study, as are the new cell volume estimates and new global diversity estimates. |
| Sampling strategy | Assemblage data are based on counts of 500 to 1000 nannofossil liths across a minimum of 10 consecutive fields of view. New data presented herein follow identical count and taxonomic protocols as the data published by Gibbs et al (2012), which span the PETM and hyperthermals. Initial observation of new samples presented herein indicated that sample size of 500 - 800 nannofossil liths was appropriate, and enabled continuity with the previously published data of Gibbs et al. (2012) and Bown (2005). |
| Data collection | New assemblage data were collected by SA using an Olympus BX43 microscope. For the immediate post-K/Pg samples, specimens were identified using an Olympus BX51 microscope in phase contrast due to the small size and low birefringence of specimens. Smear slides were placed on the microscope stage and a slide transect was selected at random. Start and end points of all transects were noted as microscope-specific slide co-ordinates to enable re-view. All nannofossil liths within each field of view were identified, and the count was continued until either a minimum of 500 specimens or 10 consecutive fields of view were identified. Data were entered initially onto hard copy count sheets and subsequently input into spreadsheets. Taxonomic consistency was verified as necessary, through discussion with PB, SG and HK. |
| Timing and spatial scale | Assemblage counts were performed between June 2015 and June 2017. Once taxonomic concepts were established, samples were typically analysed at a rate of ~5 samples per day. Data collection was not continuous, but was not typically paused for more than 30 days at a time. The immediate post-K/Pg interval was counted last, from ~January to June 2017, with Neobiscutum and Praeprinsius counts complete by June 2017. Taxonomy was checked periodically throughout, up to and including June 2018. |
| Data exclusions | Relative abundance counts from one sample (1209A 25H6 26-27 cm) were excluded due to obvious contamination with later Paleogene specimens, which were readily identifiable due to inconsistency with adjacent samples. No other samples were excluded. |
| Reproducibility | Reproducibility of findings was relatively high, specifically the reproducibility of the raw percent abundance data of the non-rare taxa. Replicates were made as standard during the course of data collection by recounting a subset of samples. In addition, replicate counts were made by SA of samples at the top and base of the section to overlap with published assemblage counts of SG and PB. We did not do any statistical tests to check reproducibility. However, the sheer volume and high-resolution of the samples supports reproducibility, and the smoothing of the trend helps to reduce the impact of any inconsistency. |
| Randomization | Samples were mostly analysed in stratigraphic order, from oldest (deepest) to youngest (shallowest). This was specifically to aid taxonomic consistency and to observe patterns of evolution as new taxa emerged. Specimens counted were selected at random by taking a random transect across the approximate middle of each smear slide, observing consecutive fields of view. The majority of slides had an even distribution of nannofossil liths but where numbers were very low (<10) or prohibitively high (such that identification was impossible) a field of view may have been skipped. Slide co-ordinates were noted to enable re-view. |
| Blinding | Blinding was not used during the data acquisition and analysis, for the reasons outlined in the 'randomization' section. We typically perform assemblage counts in stratigraphic order to aid our taxonomic consistency as it allows us to more easily identify first appearances of taxa which are often rare and can exhibit intermediate forms, as well as last occurences where again taxa can often become very rare at the top of their ranges. |

Did the study involve field work?     ☐ Yes     ☒ No

# Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

## Materials & experimental systems

| n/a | Involved in the study |
|-----|------------------------|
| ☒ | Antibodies |
| ☒ | Eukaryotic cell lines |
| ☐ ☒ | Palaeontology |
| ☒ | Animals and other organisms |
| ☒ | Human research participants |
| ☒ | Clinical data |

## Methods

| n/a | Involved in the study |
|-----|------------------------|
| ☒ | ChIP-seq |
| ☒ | Flow cytometry |
| ☒ | MRI-based neuroimaging |

## Palaeontology

Specimen provenance
All samples were provided by IODP. For new data collected at ODP Site 1209, samples were provided in Feb. 2015 under sample request 022214IODP. Additional samples at ODP Site 1209 are from Gibbs et al (2012), ODP Site 1210 are from Bown et al (2005).

Specimen deposition
Samples available from IODP.

Dating methods
The depth of each sample at ODP Site 1209 was calculated according to Westerhold and Rohl (2006) and dated according to the age models of Dinares-Turell et al (2014) and Westerhold et al (2008, option 2), updated according to Westerhold et al (2018). Samples at ODP Site 1210 were aligned to equivalent depths at ODP Site 1209 following Westerhold and Rohl (2006) and aged accordingly. See Data Table associated with Figure 1.

☒ Tick this box to confirm that the raw and calibrated dates are available in the paper or in Supplementary Information.

# LETTER

# Milk of ruminants in ceramic baby bottles from prehistoric child graves

J. Dunne[1]*, K. Rebay–Salisbury[2], R. B. Salisbury[2], A. Frisch[3], C. Walton–Doyle[1] & R. P. Evershed[1]*

**The study of childhood diet, including breastfeeding and weaning, has important implications for our understanding of infant mortality and fertility in past societies[1]. Stable isotope analyses of nitrogen from bone collagen and dentine samples of infants have provided information on the timing of weaning[2]; however, little is known about which foods were consumed by infants in prehistory. The earliest known clay vessels that were possibly used for feeding infants appear in Neolithic Europe, and become more common throughout the Bronze and Iron Ages. However, these vessels—which include a spout through which liquid could be poured—have also been suggested to be feeding vessels for the sick or infirm[3,4]. Here we report evidence for the foods that were contained in such vessels, based on analyses of the lipid 'fingerprints' and the compound-specific $\delta^{13}C$ and $\Delta^{13}C$ values of the major fatty acids of residues from three small, spouted vessels that were found in Bronze and Iron Age graves of infants in Bavaria. The results suggest that the vessels were used to feed infants with milk products derived from ruminants. This evidence of the foodstuffs that were used to either feed or wean prehistoric infants confirms the importance of milk from domesticated animals for these early communities, and provides information on the infant-feeding behaviours that were practised by prehistoric human groups.**

The study of past infancy—including infant care, breastfeeding and weaning practices—provides valuable information on population demographics and health, reproduction rates, mortality patterns and fertility of individuals of past societies. Today, feeding practices for babies can be attributed to various ecological and socioeconomic constraints and cultural factors, such as health beliefs and food taboos[1,5,6]. Prehistoric humans probably practised a range of infant-feeding behaviours[2–4,6,7], which had profound consequences for the biological and social wellbeing of the infants. Ethnographic, historical and social studies have shown differences across the breastfeeding phase, the nature of the addition of supplementary foods (during weaning) and the timing of cessation of breastfeeding[1,5,6,8].

Breastfeeding is integral to infant care in all human groups and fundamental to the mother–infant relationship[4]. Breast milk provides an infant with all of the macro- and micronutrients that are required to sustain growth for the first six months of life[9], together with bioactive components, which protect the infant from pathogenic organisms and facilitate the development and maturation of the immune system[10]. The introduction of energy and nutrient-rich, easily digestible, supplementary foods in infant feeding (that is, during weaning) is unique to humans[11,12]. Supplementary foods are generally introduced at around six months of age, when the metabolic requirements of an infant exceed the energy yield that the mother can provide through milk, contributing to the infant diet as chewing, tasting and digestive competencies develop[1,12,13].

Considerable variation exists in the practice and duration of breastfeeding and the subsequent addition of supplementary and/or weaning foodstuffs between human groups. Hunter-gatherers typically breastfeed for several years, whereas the adoption of a sedentary lifestyle in early farming communities led to shortening of the breastfeeding period[14], which was probably due to the introduction of agriculture,

at which time new foods became available to wean infants—for example, animal milk and cereal products. The widespread use of animal milk, either to feed babies or as a supplementary weaning food source, became possible with the domestication of dairy animals during the European Neolithic[14], during which time generally improved nutrition contributed to an increased birth rate, with shorter inter-birth intervals, that resulted in considerable growth of the human population: the so-called Neolithic demographic transition[15]. Broad trends identified from the Neolithic to Iron Age in Central Europe suggest that supplementary foods were given to babies at around six months of age and weaning was complete by two to three years of age[3].

Possible infant-feeding vessels that are made from clay first appear in Neolithic Europe. One of the earliest of such finds is a Linear Pottery Culture feeding vessel from Steigra, Germany, that has been dated[16] to around 5500–4800 BC. These unique vessels, which have a small spout through which liquid could be poured or suckled, come in many forms and sizes and occasionally have a zoomorphic design (Extended Data Fig. 1). They become more common in Central Europe during the late Bronze and early Iron Age[4] and are found in settlements, as stray finds, and in graves (particularly those of children), which strongly suggests that they were feeding or weaning vessels for infants.

The precious nature and often small openings of these vessels makes their sampling for organic residue analysis extremely challenging. However, infant-feeding vessels that have an open, bowl form, found in
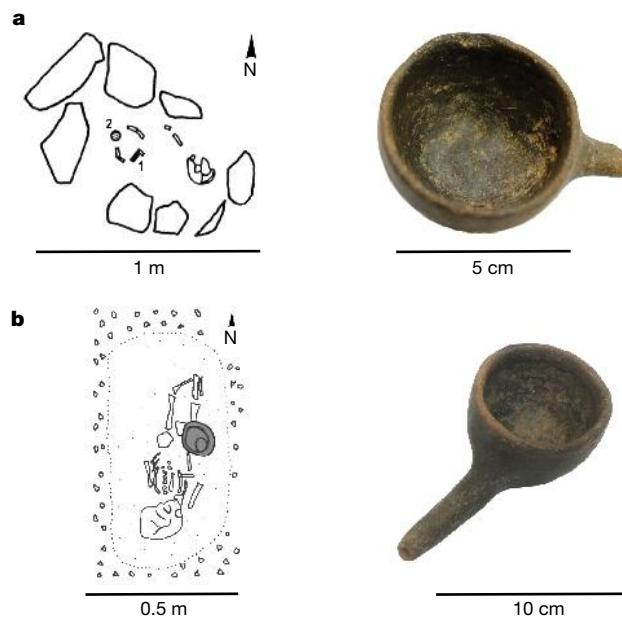


**Fig. 1 | Description of the child graves and associated feeding vessels. a**, **b**, Drawings of child graves from Dietfurt (left) and images of the feeding vessels found in each grave (right). Photographs of vessels were taken by A.F. (**a**) and K.R.-S. (**b**). Drawings of the graves were reproduced from a previously published plan[17] (**a**) and drawing[18] (**b**).

[1]Organic Geochemistry Unit, School of Chemistry, University of Bristol, Bristol, UK. [2]Institute for Oriental and European Archaeology, Austrian Academy of Sciences, Vienna, Austria. [3]Abteilung Archäologie, Museen der Stadt Regensburg, Regensburg, Germany. *e-mail: julie.dunne@bristol.ac.uk; r.p.evershed@bristol.ac.uk
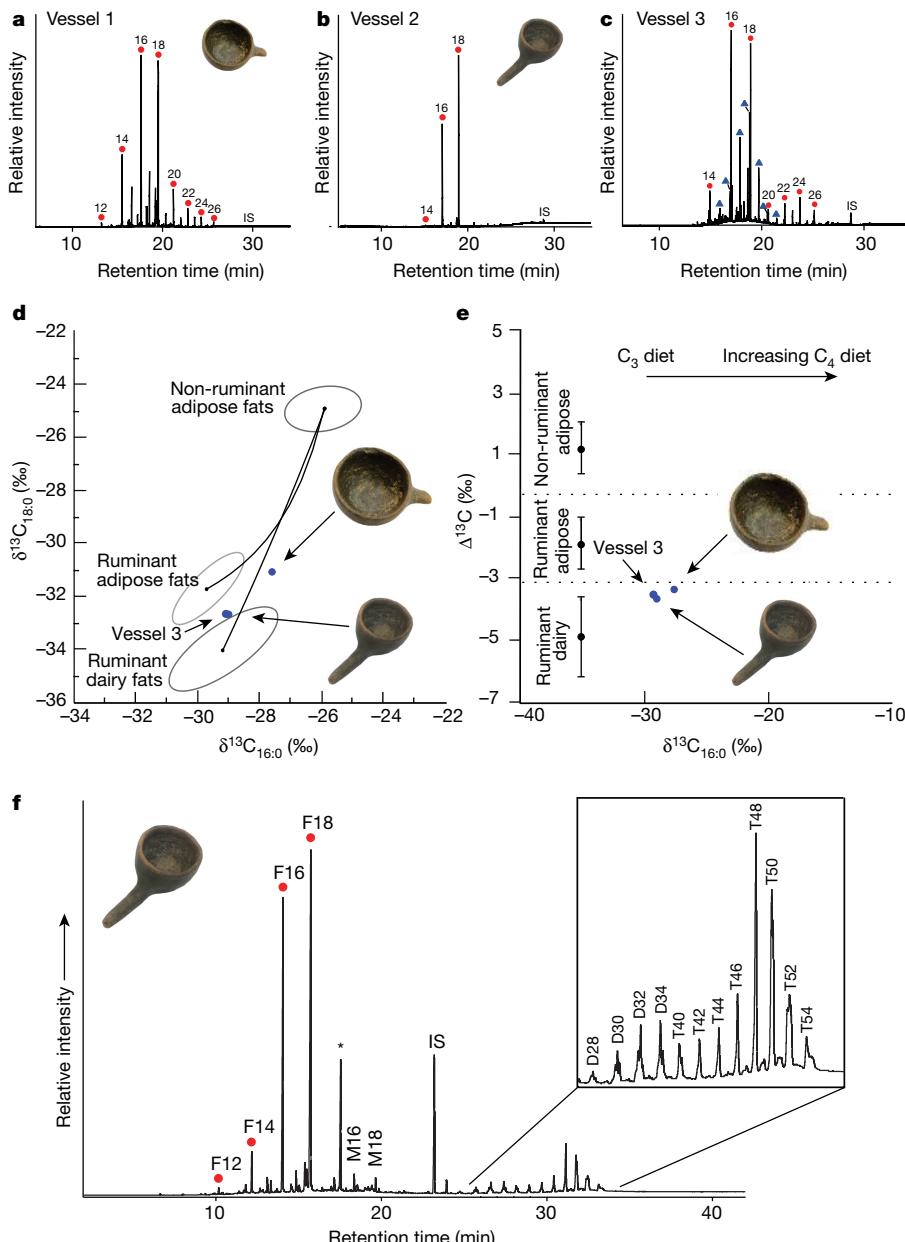
**Fig. 2 | Partial gas chromatograms and plots of δ¹³C and Δ¹³C values of *n*-alkanoic acids in infant-feeding vessels from Dietfurt and Augsburg cemeteries, Bavaria.** $n$ = 3 vessels. **a–c**, Partial gas chromatograms of transmethylated trimethylsilylated extracts from infant-feeding vessels 1–3. Red circles, *n*-alkanoic acids (fatty acids); blue triangles, *n*-alkanes; IS, internal standard, $C_{34}$ *n*-tetratriacontane. **d**, δ¹³C values for the $C_{16:0}$ and $C_{18:0}$ fatty acids for archaeological fats extracted from infant-feeding vessels 1–3. The three fields correspond to the $P = 0.684$ confidence ellipses for animals raised on a strict $C_3$ diet in Britain[20]. Each data point represents an individual vessel. **e**, The Δ¹³C ($δ^{13}C_{18:0} - δ^{13}C_{16:0}$) values are from the same vessels as in **d**. The ranges shown here represent the mean ± 1 s.d. of the Δ¹³C values from a global database comprising modern reference animal fats, which have been published previously[24]. **f**, Partial high-temperature gas chromatogram of trimethylsilylated total lipid extract of infant-feeding vessel 2, showing degraded animal fat. Red circles indicate short- and long-chain *n*-alkanoic acids with the indicated number of carbon atoms; monoacylglycerols (M) containing 16 and 18 acyl carbon atoms; diacylglycerols (D) containing 28, 30, 32 and 34 acyl carbon atoms; triacylglycerols (T), containing 40, 42, 44, 46, 48, 50, 52 and 54 acyl carbon atoms; the plasticizer is indicated by an asterisk. IS, internal standard *n*-tetratriacontane ($n$-$C_{34}$). Replication was not possible owing to the unique and irreplaceable nature of the archaeological artefacts sampled, although the objects were analysed using two different extraction methods.

graves from cemeteries of Dietfurt-Tankstelle and Dietfurt-Tennisplatz in Germany, have recently become available for chemical analysis. The graves are part of a large early Iron Age cemetery complex (dating to approximately 800–450 BC) found in the lower Altmühl valley in Bavaria, Germany, with Dietfurt-Tankstelle encompassing 99 burials in 72 graves[17] and Dietfurt-Tennisplatz containing 126 burials[18]. Child grave 80 at Dietfurt-Tennisplatz contained an east–west-oriented inhumation of a young child (0–6 years old), who had a bronze bracelet on the left arm, and in which feeding vessel 1 (Fig. 1a) was placed at the child's feet[18]. Feeding vessel 2 (Fig. 1b), which has a shape similar

to that of a small pipe, was found within a bowl deposited at the right hip in grave 65 at Dietfurt-Tankstelle, which contained a south–north-oriented inhumation of an approximately 1-year-old child[17]. Both vessels were of similar size (approximately 50 mm in diameter), although vessel 1 has a much shorter spout. A further broken vessel (vessel 3)—found in the cremation burial of a 1–2-year-old child at Augsburg-Haunstetten 1, Bavaria[19], from a late Bronze Age necropolis (around 1200–800 BC)—was also investigated.

Organic residue analyses were performed as described in previous publications[20,21], although a modified sampling procedure was adopted

for vessels 1 and 2 to minimize damage to the artefacts. Appreciable lipid content (29.7, 1.5 and 0.9 mg g$^{-1}$) was recovered from vessels 1, 2 and 3, respectively, using acidified methanol extraction, suggesting that the vessels were used for sustained processing and/or consumption of high-lipid-containing commodities. All extracts were dominated by palmitic ($C_{16}$) and stearic ($C_{18}$) fatty acids, which are typical of degraded animal fat[22] (Fig. 2a–c). Shorter-chain fatty acids ($C_{12}$ and $C_{14}$ in vessel 1 and $C_{14}$ in vessels 2 and 3) (Fig. 2a–c), which are rarely detected in archaeological pottery, were also present. The latter are likely to be remnants of $C_4$–$C_{14}$ fatty acids that are only seen in fresh milk fats[22,23], as the shorter chain homologues (shorter or equal to $C_{10}$) have been shown in degradation experiments to be lost through leaching or volatilization[20].

Further characterization of the fats was achieved through stable carbon isotope ($\delta^{13}C$) values of the $C_{16:0}$ and $C_{18:0}$ fatty acids[20,22,24]. Vessels 2 and 3 plot just outside the reference range of $\pm 1$ s.d. for dairy products, which suggests that these were primarily used to process ruminant dairy products, whereas vessel 1 (Fig. 2d) plots between the dairy and non-ruminant adipose ranges, indicating minor mixing of non-ruminant (pig or, possibly, human milk) and dairy products. The $\Delta^{13}C$ values obtained for the lipid residues from vessels 1, 2 and 3 (which were $-3.4$, $-3.7$ and $-3.6$ ‰, respectively) plot in the ruminant dairy region, consistent with the processing and/or feeding of predominantly dairy products in these vessels[24] (Fig. 2e). We interpret the results from vessel 3 with caution, owing to the presence of minor—possibly contaminating—compounds; however, as this vessel was from a cremation grave, these could be pyrolytically derived if the baby bottle was included in the funeral pyre.

As the $\Delta^{13}C$ values are found to be at the top of the range for dairy fats, the vessels were also analysed by solvent extraction[20] using high-temperature gas chromatography and high-temperature gas chromatography–mass spectrometry for diagnostic intact acyl lipids[22]. Figure 2f shows that triacylglycerols (TAGs) and their degradation products, di- and monoacylglycerols, were present in vessel 2, with TAGs comprising $C_{40}$–$C_{54}$ acyl carbon atoms with $C_{48}$ being the most abundant homologue. The latter TAGs were not detectable in vessels 1 and 3, indicating complete diagenetic hydrolysis of the acyl lipids in these vessels. Fresh adipose fats are characterized by TAGs that contain 48–54 acyl carbon atoms, whereas dairy fats are distinguished by TAGs that contain 24–54 acyl carbon atoms[23]. Whereas shorter-chain TAGs (24–38 acyl carbon atoms) are rarely seen in degraded archaeological fats, owing to diagenetic loss (which has been demonstrated experimentally[20]), $C_{40}$–$C_{46}$ TAGs are highly diagnostic of dairy fats[20,22]. In summary, our findings provide unequivocal evidence that all three vessels were predominantly used to process dairy fats.

The finding of these three obviously specialized vessels in child graves combined with our chemical evidence strongly points to these vessels having been used to feed animal milk to babies (instead of human milk) and/or children during weaning to supplementary foods (Extended Data Fig. 2). Although milk from ruminant animals may have provided a valuable extra source of nutrition, it is important to note its potential negative effect on infant health[14]. Milks are species-specific and there are key differences in the composition of human and ruminant milk. Animal milk could have been used as a supplementary food, but it would not have been a full replacement for human milk, which contains similar amounts of lipid but more carbohydrates (in the form of lactose) and considerably less protein. These differences might affect an infant in various ways. For instance, cow's milk is more difficult for an infant to absorb as it contains higher quantities of saturated fatty acids and much larger fat globules than human milk[25], causing a reduced energetic input for the infant. The processing of animal milk and the possible incorporation of meat-based gruel may have served to balance out nutritional deficiencies. However, the introduction of inappropriate supplementary foods would have provided an opportunity for infectious agents and pathogens, causing diarrhoea and other diseases, and putting the infant at greater risk of iron-deficiency anaemia[14]. These supplementary foods may also have been nutritionally inadequate, leading to malnutrition, which is detrimental to future development. Furthermore, feeding unpasteurized animal milk comes with a risk of contamination and transmission of zoonoses[4] and bacterial contamination from the vessel itself is also possible. Notwithstanding these obvious risks, our discovery of ruminant-milk-based foods in these prehistoric baby bottles offers a rare glimpse into the ways that prehistoric families were attempting to deal with the challenges of infant nutrition and weaning at this inherently risky phase of the human lifecycle.

## Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-019-1572-x.

1.  Sellen, D. W. Evolution of infant and young child feeding: implications for contemporary public health. *Annu. Rev. Nutr.* **27**, 123–148 (2007).
2.  Jay, M. Breastfeeding and weaning behaviour in archaeological populations: evidence from the isotopic analysis of skeletal materials. *Childhood in the Past* **2**, 163–178 (2009).
3.  Fulminante, F. Infant feeding practices in Europe and the Mediterranean from prehistory to the Middle Ages: a comparison between the historical sources and bioarchaeology. *Childhood in the Past* **8**, 24–47 (2015).
4.  Rebay-Salisbury, K. Breast is best – and are there alternatives? Feeding babies and young children in prehistoric Europe. *Mitt. Anthropol. Ges. Wien* **147**, 13–29 (2017).
5.  Sellen, D. W. & Smay, D. B. Relationship between subsistence and age at weaning in "preindustrial" societies. *Hum. Nat.* **12**, 47–87 (2001).
6.  Halcrow, S. E. et al. in *Breastfeeding : New Anthropological Approaches* (ed. Tomori, C. et al.) 155–169 (Routledge, 2017).
7.  Mays, S., Gowland, R., Halcrow, S. & Murphy, E. Child bioarchaeology: perspectives on the past 10 years. *Childhood in the Past* **10**, 38–56 (2017).
8.  Fildes, V. *Breasts, Bottles and Babies: A History of Infant Feeding* (Edinburgh Univ. Press, 1986).
9.  Kramer, M. S. & Kakuma, R. Optimal duration of exclusive breastfeeding. *Cochrane Database Syst. Rev.* **8**, CD003517 (2012).
10. Robinson, S. M. Infant nutrition and lifelong health: current perspectives and future challenges. *J. Dev. Orig. Health Dis.* **6**, 384–389 (2015).
11. Knott, C. D. in *Reproductive Ecology and Human Evolution* (ed. Ellison, P. T.) 429–463 (Aldine de Gruyter, 2001).
12. Humphrey, L. T. Weaning behaviour in human evolution. *Semin. Cell Dev. Biol.* **21**, 453–461 (2010).
13. Sheppard, J. J. & Mysak, E. D. Ontogeny of infantile oral reflexes and emerging chewing. *Child Dev.* **55**, 831–843 (1984).
14. Howcroft, R., Eriksson, G. & Lidén, K. The Milky Way: the implications of using animal milk products in infant feeding. *Anthropozoologica* **47**, 31–43 (2012).
15. Bocquet-Appel, J. P. When the world's population took off: the springboard of the Neolithic demographic transition. *Science* **333**, 560–561 (2011).
16. Meller, H. E. *Bronzerausch: Spätneolithikum und Frühbronzezeit. Begleithefte zur Dauerausstellung 4* (Landesamt für Denkmalpflege und Archäologie Sachsen-Anhalt, Landesmuseum für Vorgeschichte Halle, 2011).
17. Augstein, M. *Das Gräberfeld der Hallstatt- und Frühlatènezeit von Dietfurt an der Altmühl (›Tankstelle‹). Ein Beitrag zur Analyse einer Mikroregion Universitätsforschungen zur Prähistorischen Archäologie 262* (Habelt, 2015).
18. Röhrig, K.-H. *Das hallstattzeitliche Gräberfeld von Dietfurt a.d. Altmühl* (Leidorf, 1994).
19. Wirth, S. *Grabfunde der späten Bronzezeit und der Urnenfelderzeit von Augsburg-Haunstetten und von Friedberg in Bayern. Ein Beitrag zur vorgeschichtlichen Besiedlung des unteren Lechtals Augsburger Beiträge zur Archäologie 1* (Wissner, 1998).
20. Dudd, S. N. & Evershed, R. P. Direct demonstration of milk as an element of archaeological economies. *Science* **282**, 1478–1481 (1998).
21. Correa-Ascencio, M. & Evershed, R. P. High throughput screening of organic residues in archaeological potsherds using direct acidified methanol extraction. *Anal. Methods* **6**, 1330–1340 (2014).
22. Evershed, R. P. et al. Chemistry of archaeological animal fats. *Acc. Chem. Res.* **35**, 660–668 (2002).
23. Fontecha, J., Goudjil, H., Ríos, J. J., Fraga, M. J. & Juárez, M. Identity of the major triacylglycerols in ovine milk fat. *Int. Dairy J.* **15**, 1217–1224 (2005).
24. Dunne, J. et al. First dairying in green Saharan Africa in the fifth millennium bc. *Nature* **486**, 390–394 (2012).
25. Gurr, M. I. Review of the progress of dairy science: human and artificial milks for infant feeding. *J. Dairy Res.* **48**, 519–554 (1981).

## METHODS

Lipid analyses were performed mostly as described in previous publications[20,21], except that a modified sampling procedure was adopted for vessels 1 and 2 to minimize damage to the artefacts. Part of the internal surface layer of each vessel was removed by abrasion to remove contamination, after which the underlying material was taken as a powder for analysis of absorbed organic residues (approximately 0.84 g and 0.33 g for vessels 1 and 2, respectively). A small fragment (approximately 0.95 g) of vessel 3 was destructively sampled after surface cleaning by abrasion and grinding to powder. All solvents used were HPLC grade (Rathburn) and the reagents were analytical grade (typically >98% of purity). An internal standard (typically 40 μg) was added to enable quantification of the lipid extract (n-tetratriacontane; Sigma-Aldrich). Following the addition of 5 ml $H_2SO_4$/methanol 2–4% ($\delta^{13}C$ measured), the culture tubes were placed on a heating block for 1 h at 70 °C, mixing every 10 min. Once cooled, the methanolic acid was transferred to test tubes and centrifuged at 2,500 r.p.m. for 10 min. The supernatant was then decanted into another furnaced culture tube and 2 ml dichloromethane-extracted double-distilled water was added. To recover any lipids not fully solubilized by the methanol solution, $2 \times 3$ ml of n-hexane was added to the extracted potsherds contained in the original culture tubes, mixed well and transferred to the second culture tube. The extraction was transferred to a clean, furnaced 3.5-ml vial and blown down to dryness. Following this, $2 \times 2$ ml n-hexane was added directly to the $H_2SO_4$/methanol solution in the second culture tube and whirlimixed to extract the remaining residues. This was transferred to the 3.5-ml vials and blown down under a gentle stream of nitrogen until a full vial of n-hexane remained. Aliquots of the extracts (containing fatty acid methyl esters (FAMEs)) were derivatized using N,O-bis(trimethylsilyl)trifluoroacetamide (BSTFA) containing 1% v/v trimethylchlorosilane (Sigma-Aldrich; 20 μl; 70 °C, 1 h). Excess BSTFA was removed under nitrogen and the extract was dissolved in n-hexane for analysis by gas chromatography, gas chromatography–mass spectrometry and gas chromatography–combustion–isotope ratio mass spectrometry.

Further analysis was carried out using the solvent extraction method. An internal standard was added to the sherd powder and the samples were solvent-extracted by ultrasonication (chloroform:methanol 2:1 v/v, 30 min, $2 \times 10$ ml). The solvent was evaporated under a gentle stream of nitrogen to obtain the total lipid extract. Aliquots of the total lipid extract were trimethylsilylated (BSTFA, 20 μl, 70 °C, 1 h), diluted with n-hexane and analysed by high-temperature gas chromatography and high-temperature gas chromatography–mass spectrometry.

All FAMEs initially underwent high-temperature gas chromatography using a gas chromatograph fitted with a high-temperature non-polar column (DB1-HT; 100% dimethylpolysiloxane, 15 m × 0.32-mm inner diameter, 0.1-μm film thickness). The carrier gas was helium and the temperature programme comprised a 50-°C isothermal hold followed by an increase to 350 °C at a rate of 10 °C min$^{-1}$ followed by an isothermal hold at 350 °C for 10 min. A procedural blank (no sample) was prepared and analysed alongside each batch of samples. Further compound identification was accomplished using gas chromatography–mass spectrometry. FAMEs were then introduced by autosampler onto a gas chromatography–mass spectrometry setup fitted with a non-polar column (100% dimethylpolysiloxane stationary phase; 60 m × 0.25-mm inner diameter, 0.1-μm film thickness). The instrument was a ThermoFinnigan single quadrupole TraceMS that was run in electron ionization mode (electron energy 70 eV, scan time of 0.6 s). Samples were

run in full scan mode (m/z 50–650) and the temperature programme comprised an isothermal hold at 50 °C for 2 min, ramping to 300 °C at 10 °C min$^{-1}$, followed by an isothermal hold at 300 °C for 15 min. Data acquisition and processing were carried out using the HP Chemstation software (Rev. C.01.07 (27), Agilent Technologies) and Xcalibur software (v.3.0). Peaks were identified on the basis of their mass spectra and gas chromatography retention times, and compared with the NIST mass spectral library (v.2.0).

Carbon isotope analyses by gas chromatography–combustion–isotope ratio mass spectrometry were carried out using a GC Agilent Technologies 7890A coupled to an Isoprime 100 (electron ionization at 70 eV, three Faraday cup collectors m/z 44, 45 and 46) using an IsoprimeGC5 combustion interface with a CuO and silver wool reactor maintained at 850 °C. Instrument accuracy was determined using an external FAME standard mixture ($C_{11}$, $C_{13}$, $C_{16}$, $C_{21}$ and $C_{23}$) of known isotopic composition. Samples were run in duplicate and an average taken. The $\delta^{13}C$ values are the ratios $^{13}C/^{12}C$ and expressed relative to the Vienna Pee Dee Belemnite, calibrated against a $CO_2$ reference gas of known isotopic composition. Instrument error was ±0.3‰. Data processing was carried out using Ion Vantage software (v.1.6.1.0, IsoPrime).

**Reporting summary.** Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

## Data availability

All data generated or analysed during this study are included in this published Letter.

**Extended Data Fig. 1 | Selection of late Bronze/early Iron Age feeding vessels.** Vessels are from Vienna, Oberleis, Vösendorf and Franzhausen-Kokoron (from left to right), dated to around 1200–800 BC. The vessels are approximately 105, 80, 90 and 80 mm high (from left to right). Photographs were taken by K.R.-S.

**Extended Data Fig. 2 | Modern-day baby feeding from reconstructed infant-feeding vessel of the type investigated in this study.** Photograph was taken by H. Seidl da Fonseca.

# natur:research

Corresponding author(s):   Richard P. Evershed

# Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

## Statistical parameters

When statistical analyses are reported, confirm that the following items are present in the relevant location (e.g. figure legend, table legend, main text, or Methods section).

| n/a | Confirmed | |
|---|---|---|
| ☒ | ☐ | The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement |
| ☐ | ☒ | An indication of whether measurements were taken from distinct samples or whether the same sample was measured repeatedly |
| ☒ | ☐ | The statistical test(s) used AND whether they are one- or two-sided<br>*Only common tests should be described solely by name; describe more complex techniques in the Methods section.* |
| ☒ | ☐ | A description of all covariates tested |
| ☒ | ☐ | A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons |
| ☒ | ☐ | A full description of the statistics including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals) |
| ☒ | ☐ | For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted<br>*Give P values as exact values whenever suitable.* |
| ☒ | ☐ | For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings |
| ☒ | ☐ | For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes |
| ☒ | ☐ | Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated |
| ☐ | ☒ | Clearly defined error bars<br>*State explicitly what error bars represent (e.g. SD, SE, CI)* |

*Our web collection on statistics for biologists may be useful.*

## Software and code

Policy information about availability of computer code

| Data collection | No customised computer code was used in the collection of the data presented in this paper. Gas chromatography data were acquired using HP Chemstation software (Rev. C.01.07 (27). GC/MS data were acquired using XCalibur software (version 3). GC-C-IRMS were acquired using Ion Vantage software (version 1.6.1.0, IsoPrime). |
|---|---|
| Data analysis | No customised computer code was used in the processing of the data presented in this paper. Gas chromatography data were processed using HP Chemstation software (Rev. C.01.07 (27). GC/MS data were processed using XCalibur software (version 3). GC-C-IRMS were processed using Ion Vantage software (version 1.6.1.0, IsoPrime). |

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

## Data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

> All new data is provided in the paper. Reference data is taken from previously published papers cited in the manuscript.

# Field-specific reporting

Please select the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences          ☐ Behavioural & social sciences          ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/authors/policies/ReportingSummary-flat.pdf

# Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

| | |
|---|---|
| Sample size | The study focuses on unique archaeological materials wherein we analysed all examples available to us. |
| Data exclusions | There were no data exclusions. |
| Replication | Due to the unique and irreplaceable nature of the archaeological objects replication was not possible although two objects were extracted using two different extraction methods, as discussed in the text. Stable isotope analyses were performed in duplicate. |
| Randomization | Randomisation was not applicable as we were undertaking analyses of rare archaeological objects for which the numbers available were necessarily limited. |
| Blinding | Blinding was not applicable as we were undertaking analyses of rare archaeological objects for which the numbers available were necessarily limited. |

# Reporting for specific materials, systems and methods

## Materials & experimental systems

| n/a | Involved in the study |
|---|---|
| ☒ | ☐ Unique biological materials |
| ☒ | ☐ Antibodies |
| ☒ | ☐ Eukaryotic cell lines |
| ☒ | ☐ Palaeontology |
| ☒ | ☐ Animals and other organisms |
| ☒ | ☐ Human research participants |

## Methods

| n/a | Involved in the study |
|---|---|
| ☒ | ☐ ChIP-seq |
| ☒ | ☐ Flow cytometry |
| ☒ | ☐ MRI-based neuroimaging |

# LETTER

# The RIPK4–IRF6 signalling axis safeguards epidermal differentiation and barrier function

Nina Oberbeck[1], Victoria C. Pham[2], Joshua D. Webster[3], Rohit Reja[4], Christine S. Huang[5], Yue Zhang[4], Merone Roose-Girma[6], Søren Warming[6], Qingling Li[2], Andrew Birnberg[2], Weng Wong[2], Wendy Sandoval[2], László G. Kőműves[3], Kebing Yu[2], Debra L. Dugger[1], Allie Maltzman[1], Kim Newton[1] & Vishva M. Dixit[1]*

**The integrity of the mammalian epidermis depends on a balance of proliferation and differentiation in the resident population of stem cells[1]. The kinase RIPK4 and the transcription factor IRF6 are mutated in severe developmental syndromes in humans, and mice lacking these genes display epidermal hyperproliferation and soft-tissue fusions that result in neonatal lethality[2–5]. Our understanding of how these genes control epidermal differentiation is incomplete. Here we show that the role of RIPK4 in mouse development requires its kinase activity; that RIPK4 and IRF6 expressed in the epidermis regulate the same biological processes; and that the phosphorylation of IRF6 at Ser413 and Ser424 primes IRF6 for activation. Using RNA sequencing (RNA-seq), histone chromatin immunoprecipitation followed by sequencing (ChIP–seq) and assay for transposase-accessible chromatin using sequencing (ATAC-seq) of skin in wild-type and IRF6-deficient mouse embryos, we define the transcriptional programs that are regulated by IRF6 during epidermal differentiation. IRF6 was enriched at bivalent promoters, and IRF6 deficiency caused defective expression of genes that are involved in the metabolism of lipids and the formation of tight junctions. Accordingly, the lipid composition of the stratum corneum of $Irf6^{-/-}$ skin was abnormal, culminating in a severe defect in the function of the epidermal barrier. Collectively, our results explain how RIPK4 and IRF6 function to ensure the integrity of the epidermis and provide mechanistic insights into why developmental syndromes that are characterized by orofacial, skin and genital abnormalities result when this axis goes awry.**

To determine whether the kinase activity of RIPK4 is essential for its role in epidermal differentiation, we engineered $Ripk4^{D161N/D161N}$ mice that express catalytically inactive RIPK4 (Extended Data Fig. 1a). In contrast to wild-type RIPK4, which was difficult to detect in skin at embryonic day (E)16.5, RIPK4(D161N) was more abundant (Extended Data Fig. 1b). Nevertheless, $Ripk4^{D161N/D161N}$ embryos resembled $Ripk4^{-/-}$ embryos at E18.5, with thick and shiny skin (Fig. 1a) and fusion of all external orifices[6,7] (Fig. 1b). Consequently, $Ripk4^{D161N/D161N}$ pups were not viable (Extended Data Fig. 1c). Histological analysis of the skin revealed epidermal hyperplasia with proliferation of suprabasal cells, absence of the stratum corneum and parakeratosis (Fig. 1c, d). Thus, the kinase activity of RIPK4 is crucial for development and epidermal differentiation in mice.

Mutations in human *RIPK4* result in Bartsocas–Papas syndrome (BPS), a severe autosomal recessive disorder that is characterized by craniofacial, genital and skin abnormalities, including popliteal webbing and syndactyly[5,8]. Most patients with BPS die in utero or shortly after birth[9]. Mutations in human *IRF6* give rise to either van der Woude syndrome (VWS)—the most common form of syndromic cleft lip and palate[10]—or popliteal-pterygium syndrome (PPS)[4,11], which is a less severe form of BPS[12]. Mice that lack IRF6, like $Ripk4^{-/-}$ mice, display soft-tissue fusions, epidermal hyperplasia (the spinous layer is expanded but the granular layer is absent) and parakeratosis, but

they also have skeletal abnormalities that are not seen in $Ripk4^{-/-}$ or $Ripk4^{D161N/D161N}$ mice[3,13]. It has been reported that RIPK4 can phosphorylate and activate IRF6 in vitro[14,15], but the mechanism by which these proteins interact in vivo is unknown. We therefore tested whether RIPK4 and IRF6 might function in a linear pathway to control epidermal differentiation. First, we deleted conditional alleles of *Ripk4* or *Irf6* (Extended Data Fig. 1d, e) in the epidermis alone, using a K14-Cre transgene[16]. These epidermal knockouts—which we hereafter refer to as $Ripk4^{EKO}$ and $Irf6^{EKO}$ mice—lacked RIPK4 and IRF6 specifically in the epidermis (Extended Data Fig. 2a, g).

No $Ripk4^{EKO}$ or $Irf6^{EKO}$ mice were observed at clipping (postnatal day (P)4–P7) (Extended Data Fig. 2b, h). P0 pups had sticky, brittle skin (Extended Data Fig. 2c, i), and died within a few hours of birth with barrier defects at the head and extremities (Extended Data Fig. 2d, j). Unlike the total-body knockouts, $Ripk4^{EKO}$ and $Irf6^{EKO}$ mice had equivalent phenotypes (Extended Data Fig. 2e, k). $Irf6^{EKO}$ mice lacked skeletal and limb abnormalities, which indicates that the role of IRF6 in skeletal development is independent of its role in the differentiation of keratinocytes[17]. The epidermis of both mutants displayed hyperplasia, an expansion of the spinous layer, segmental loss of the granular layer and focal parakeratosis (the latter slightly more severe in the dorsum of $Ripk4^{EKO}$ embryos) (Extended Data Fig. 2e, k). These phenotypes were consistently more severe in the skin over the head than in the skin of the dorsum. Neither $Ripk4^{EKO}$ nor $Irf6^{EKO}$ mice displayed soft-tissue fusions (Extended Data Fig. 2f, l), which suggests that there is a cell-autonomous requirement for these genes in the periderm[7,17]. The results shown here for $Ripk4^{EKO}$ mice are consistent with a previously published study[18]. Collectively, our data indicate that RIPK4 and IRF6 are required in a cell-autonomous manner for keratinocyte differentiation. At least in the epidermis, RIPK4 and IRF6 appear to perform the same function(s). In accordance with this notion, $Irf6^{-/-}Ripk4^{D161N/D161N}$ embryos were indistinguishable from $Irf6^{-/-}$ embryos (Fig. 2a, b, Extended Data Fig. 3a, b), and $Irf6^{+/-}Ripk4^{+/D161N}$ double heterozygotes were weaned at Mendelian ratios (Extended Data Fig. 3c).

By RNA-seq, genes that were significantly altered in expression in E15.5 $Ripk4^{D161N/D161N}$ or $Irf6^{-/-}$ epidermis compared to wild-type epidermis ($\log_2$(gene expression in mutant/gene expression in wild type) $< -1$ or $> 1$ and adjusted $P < 0.05$ by two-sided, moderated $t$-test) were highly correlated ($R = 0.72$; Fig. 2c). In addition, all of the changes in gene expression that were seen in the $Ripk4^{D161N/D161N}$ epidermis were contained within the changes seen in the $Irf6^{-/-}$ epidermis. Thus, there may be RIPK4-dependent and RIPK4-independent ways of activating IRF6.

Phosphorylation of members of the IRF family at serine and threonine residues in the C-terminal region is known to regulate their transcriptional activity[19]. Mass spectrometry confirmed that RIPK4 phosphorylated IRF6 at Ser413 and Ser424[14] (Fig. 3a, Extended Data Fig. 4a–d), and we therefore investigated whether mutating both phosphorylation sites in IRF6 was deleterious in mice. Indeed, $Irf6^{S413A,S424A/S413A,S424A}$

[1]Department of Physiological Chemistry, Genentech, South San Francisco, CA, USA. [2]Department of Microchemistry, Proteomics and Lipidomics, Genentech, South San Francisco, CA, USA. [3]Department of Pathology, Genentech, South San Francisco, CA, USA. [4]Department of Bioinformatics and Computational Biology, Genentech, South San Francisco, CA, USA. [5]Department of Protein Chemistry, Genentech, South San Francisco, CA, USA. [6]Department of Molecular Biology, Genentech, South San Francisco, CA, USA. *e-mail: dixit@gene.com
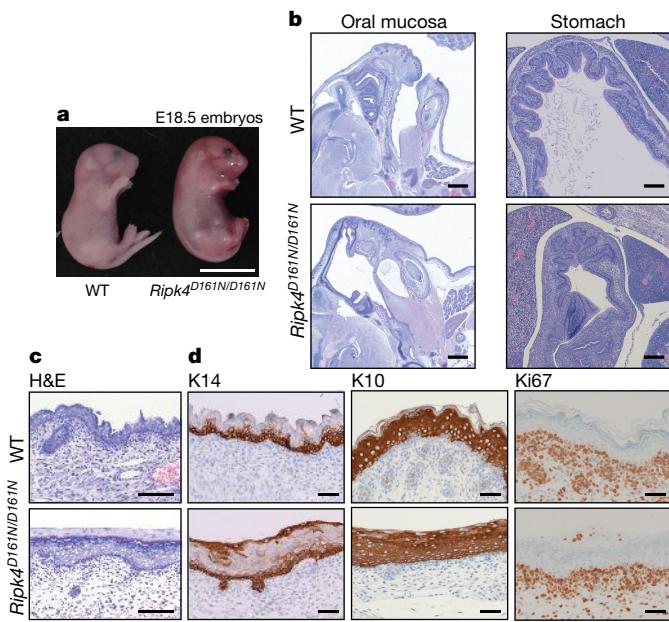
**Fig. 1 | RIPK4 kinase activity is required for epidermal differentiation and development in mice. a**, E18.5 embryos ($n = 3$ wild type (WT) and $n = 3$ $Ripk4^{D161N/D161N}$). Scale bar, 1 cm. **b**, E18.5 sections ($n = 3$ wild type and $n = 3$ $Ripk4^{D161N/D161N}$) showing fusion of the squamous epithelium at the mouth and fusion of the tongue to the palate (left), and fusion of the stratified squamous epithelial portion of the stomach (right), in $Ripk4^{D161N/D161N}$ embryos. Sections were stained with haematoxylin and eosin (H&E). Scale bars, 800 μm (left); 200 μm (right). **c**, E18.5 H&E-stained skin sections ($n = 3$ wild type and $n = 3$ $Ripk4^{D161N/D161N}$). Scale bars, 100 μm. **d**, E18.5 skin sections ($n = 3$ wild type and $n = 3$ $Ripk4^{D161N/D161N}$) immunolabelled for cytokeratin 10 (K10), K14 and Ki67 (brown). Scale bars, 50 μm.

embryos were indistinguishable from $Irf6^{-/-}$ embryos (Fig. 3b, c, Extended Data Figs. 1f, 4e, 5a–c). RNA-seq analysis demonstrated that the changes in gene expression that were seen in E15.5 $Irf6^{S413A,S424A/S413A,S424A}$ and $Irf6^{-/-}$ epidermis compared to wild-type epidermis were highly correlated ($R = 0.72$; Extended Data Fig. 4f). Thus, Ser413 and/or Ser424 of IRF6 are essential for its function in vivo.

Consistent with our genetic data, co-transfection of RIPK4 enhanced IRF6 activity in a luciferase promoter assay, and this required the kinase activity of RIPK4, as well as the IRF6 phosphorylation-site residues Ser413 and Ser424 (Extended Data 5d). Phosphomimetic mutations (Ser to Glu) at Ser413 and Ser424 of IRF6 increased the activity of IRF6 by approximately fourfold. However, RIPK4 further stimulated the activity of IRF6(S413E/S424E), which suggests that IRF6 may have additional phosphorylation sites.

If phosphorylation of IRF6 by RIPK4 is critical during development, then an IRF6(S413E/S424E) phosphomimetic knock-in mouse ($Irf6^{S413E,S424E}$) might eliminate the need for RIPK4 enzymatic activity, and thereby rescue the lethality of $Ripk4^{D161N/D161N}$ mice (Extended Data Fig. 1g). Notably, at E18.5, double-mutant $Irf6^{S413E,S424E/S413E,S424E}Ripk4^{D161N/D161N}$ embryos phenocopied $Irf6^{-/-}$ embryos. However, one wild-type allele of $Ripk4$ (that is, $Irf6^{S413E,S424E/S413E,S424E}Ripk4^{+/D161N}$) was sufficient to reverse this phenotype and give rise to a mouse with normal epidermal differentiation and skeletal development (Fig. 3d, e, Extended Data Fig. 5e–i). This result suggests that functional RIPK4—encoded by the wild-type allele—probably phosphorylates IRF6 at sites that are distinct from Ser413 and Ser424, to institute rescue.

Phosphorylation of IRF6 at Ser413 and Ser424 is clearly necessary but not sufficient to ensure its activation, which requires RIPK4-dependent phosphorylation on additional IRF6 residue(s). Stable isotope labelling by amino acids in cell culture (SILAC) and mass spectrometry analyses (Extended Data Fig. 6a) identified 10 additional phosphorylation sites in IRF6, of which Ser90 had the highest dependency on RIPK4 ($P = 2.83 \times 10^{-10}$; Fig. 3f). Indeed, the only site that resulted in a decrease in IRF6 activity when mutated to alanine was Ser90 (Fig. 3g, Extended Data Fig. 6b), which suggests that Ser90 is probably the additional IRF6 site that is required for its activation. In agreement with this mouse data, human IRF6 mutations at Ser90 (S90G) lead to VWS[20], and at Ser424 (S424L) give rise to PPS[4] (Extended Data Fig. 6c). The fact that phosphorylation at Ser424 is essential for IRF6 activation may explain why IRF6(S424L) gives rise to a developmental syndrome even though it is one of only two PPS-associated mutations that are located outside of the DNA-binding domain of IRF6[4].

An in vitro kinase assay confirmed that the RIPK4 kinase domain (but not a catalytically inactive version) could directly phosphorylate IRF6 at Ser413, Ser424 and Ser90 (Extended Data Fig. 6d). In addition, SILAC and mass spectrometry analyses showed that mutating IRF6 Ser413 and Ser424 to alanine reduced the amount of phosphorylated



**Fig. 2 | IRF6 and RIPK4 lie on a linear pathway in epidermal differentiation. a**, E18.5 embryos ($n = 3$ each for wild type, $Ripk4^{D161N/D161N}$, $Irf6^{-/-}$ and $Ripk4^{D161N/D161N}Irf6^{-/-}$). Scale bar, 1 cm. **b**, E18.5 H&E-stained skin sections from embryos with the indicated genotypes ($n = 3$ each). Scale bars, 100 μm. **c**, Genome-wide 'four-way' plot showing genes that have increased or decreased expression in E15.5 skin from $Ripk4^{D161N/D161N}$ ($y$ axis; $n = 5$) or $Irf6^{-/-}$ ($x$ axis; $n = 5$) embryos versus wild type ($n = 3$). Each coloured dot represents a gene that met the cut-offs of an

adjusted $P$ value $< 0.05$ and a minimum twofold change in expression. Adjusted $P$ values were obtained using a moderated $t$-test (two-sided) with the Benjamini–Hochberg method for multiple comparisons. Genes that were altered significantly in expression in $Irf6^{-/-}$ only are shown in red, those altered in $Ripk4^{D161N/D161N}$ only are shown in green and those altered in both genotypes are shown in blue. The Pearson correlation coefficient ($R$ value) is 0.72.

**Fig. 3 | Phosphorylation of IRF6 at Ser413 and Ser424 primes IRF6 for activation, and is essential for epidermal differentiation and development. a**, Quantification of phosphorylation at Ser413 or Ser424 of IRF6 co-expressed with either wild-type or kinase-dead (KD) RIPK4 (RIPK4(D161N)) (two independent experiments). Data are the mean ($\pm$s.d.) percentage of sites that are phosphorylated. p, phosphorylated residue. **b**, E18.5 embryos ($n = 3$ wild type and $n = 3$ $Irf6^{S413A,S424A/S413A,S424A}$). Scale bar, 1 cm. **c**, E18.5 H&E-stained skin sections ($n = 3$ wild type and $n = 3$ $Irf6^{S413A,S424A/S413A,S424A}$). Scale bars, 50 μm. **d**, E18.5 embryos of the indicated genotypes ($n = 3$ each). Scale bar, 1 cm. **e**, E18.5 H&E-stained skin sections of the indicated genotypes ($n = 3$ each). Scale bars, 50 μm. **f**, Schematic of RIPK4-dependent phosphorylation sites on IRF6 (blue dots). Blue–yellow colour scale indicates $\log_2$(RIPK4(D161N)/ wild-type RIPK4). IRF6 Y253 and Y322 are black ($\log_2$ ratio around 0) because RIPK4 is a serine/threonine kinase. Dots connected by lines indicate doubly phosphorylated sites covered by one peptide. Bars above the protein sequence represent non-phosphorylated peptides from trypsin and chymotrypsin digestions (two experimental replicates). Most lines are black ($\log_2$ ratio close to 0), indicating that similar amounts of each peptide were present with wild-type RIPK4 or RIPK4(D161N). Peptides in the S413 and S424 region are yellow, indicating that less unmodified peptide was present with wild-type RIPK4 than RIPK4(D161N). Dotted lines indicate $P$ values (lme function, two-sided, not adjusted for multiple comparisons; calculated from two independent experiments); the further down the phosphorylation site extends, the higher the confidence in its identification. $P < 0.001$ for pS413, pS424 and pS90. **g**, Graph indicates activation of an IRF-responsive luciferase reporter gene after transfection of 293T cells with the indicated IRF6 and RIPK4 constructs. Mean IRF6 activity ($\pm$ s.d) is displayed as fold activity over reporter only ($n = 4$ independent experiments). Unpaired, two-tailed Fisher's exact test with 95% confidence interval.

at K27 (H3K27ac; associated with active enhancers)) on wild-type and $Irf6^{-/-}$ E16.5 whole skin to identify putative IRF6 target genes (Extended Data Fig. 8a).

RNA-seq analysis showed that 1,226 genes were significantly downregulated and 458 significantly upregulated in $Irf6^{-/-}$ compared to wild-type skin ($\log_2$(gene expression in $Irf6^{-/-}$/gene expression in wild type) $< -1$ or $> 1$ and adjusted $P < 0.05$ by two-sided, moderated $t$-test). Gene-set enrichment analysis on the downregulated genes revealed that lipid metabolism pathways were perturbed. Tight junctions also appeared on the list (Fig. 4b). These data are relevant mechanistically; the extracellular lipid matrix of the stratum corneum and tight junctions are two indispensable components of the epidermal barrier that are essential for survival, and $Irf6^{-/-}$ (and $Ripk4^{-/-}$) embryos have a severe defect in the epidermal barrier[7] (Extended Data Fig. 8d, e).

We examined our histone ChIP–seq and ATAC-seq data to determine whether any genes involved in lipid metabolism or tight junctions were targets of IRF6. IRF6 has been shown to recognize and bind to interferon-stimulated response element (ISRE) sites[23]. We were able to detect the ISRE motif only in our H3K4me3 ChIP–seq dataset (Fig. 4c), which indicates that IRF6-binding sites may be enriched at promoters. By combining the IRF6 motif data with ATAC-seq and RNA-seq data, we identified a list of 66 high-confidence IRF6 targets (Extended Data Fig. 8a–c). In $Irf6^{-/-}$ skin, 45 of these genes were downregulated, suggesting that IRF6 may activate their transcription. Notably, compared to background clustering of all genes, putative IRF6 targets were significantly enriched at both active and bivalent promoters ($P = 8.8 \times 10^{-16}$ and $P = 2.4712 \times 10^{-11}$, respectively; two sample $z$-test for proportions). This 'poised' state at bivalent promoters is common for genes that are involved in development, as it allows a timely response to differentiation signals[24] (Extended Data Fig. 8c).

One of these 45 downregulated genes, grainyhead-like transcription factor 3 (*Grhl3*), is mutated in patients with VWS, and *Grhl3*$^{-/-}$ mice exhibit altered lipid processing and die at birth owing to a defect in the epidermal barrier[25–27] (Extended Data Fig. 8h, 9a). Another gene, occludin (*Ocln*), encodes a component of tight junctions[28] (Extended Data Figs. 8f, 9a). Four genes are involved in the biosynthesis of stratum

Ser90 by half compared to wild-type IRF6, whereas mutating IRF6 Ser90 to alanine had no effect on phosphorylation at Ser413 and Ser424 (Extended Data Fig. 7a–e). Phosphomimetic mutations at Ser413 and Ser424 stimulated a 1.5-fold increase in phosphorylation at Ser90 (Extended Data Fig. 7f). This suggests a two-step phosphorylation mechanism in which phosphorylation of Ser413 and Ser424 enhances phosphorylation at Ser90 (but not the other way around). These genetic and biochemical results allowed us to generate a model to explain IRF6 activation in vivo (Extended Data Fig. 6e).

Although our data explain how RIPK4 and IRF6 function in concert to control epidermal differentiation, the in vivo transcriptional targets of IRF6 that are responsible for epidermal differentiation remain unknown[21,22]. Immunohistochemistry (IHC) of IRF6 showed that IRF6 is expressed in all layers of the epidermis apart from the stratum corneum (Fig. 4a). We therefore performed RNA-seq, ATAC-seq and ChIP–seq for three histone modifications (histone H3 trimethylated at K4 (H3K4me3; associated with active promoters) or K27 (H3K27me3; associated with repressed chromatin), and H3 acetylated
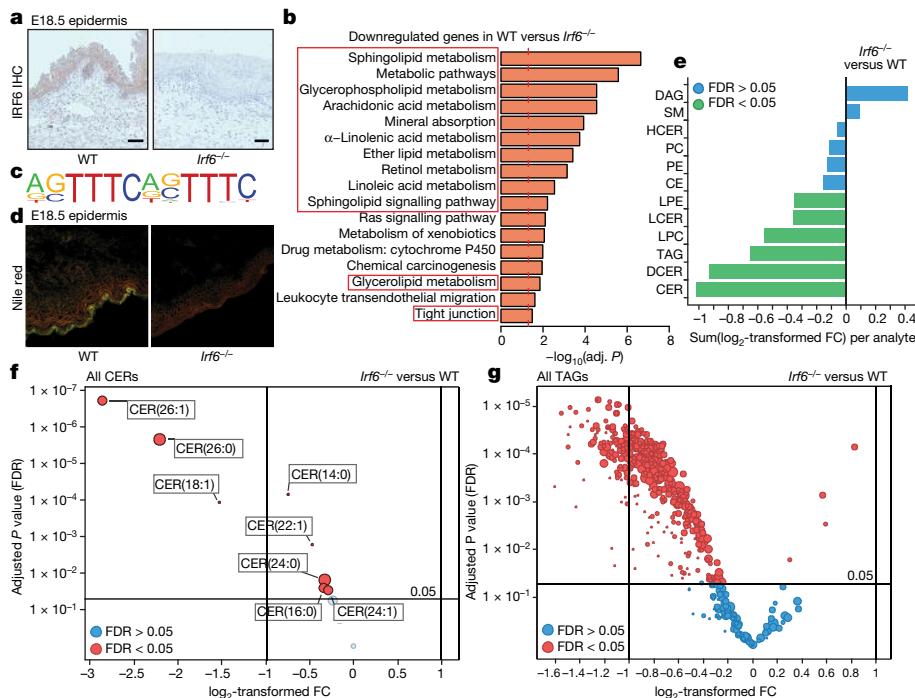
**Fig. 4 | IRF6 regulates the expression of genes that are involved in lipid metabolism, and is essential for normal metabolism of epidermal lipids and barrier function. a,** E18.5 skin sections ($n = 3$ wild type and $n = 3$ $Irf6^{-/-}$) immunolabelled for IRF6 (brown). Scale bars, 50 μm. **b,** KEGG pathway enrichment analysis on genes that were significantly downregulated (log2(gene expression in $Irf6^{-/-}$/gene expression in wild type) $< -1$, adjusted $P$ value $< 0.05$) in $Irf6^{-/-}$ compared to wild-type E16.5 skin ($n = 3$ each) reveals genes involved in lipid metabolism and tight-junction formation (red boxes). Hypergeometric test (one-sided) with Benjamini–Hochberg correction. **c,** IRF6 motif (ISRE) identified by performing de novo motif analysis on H3K4me3 peaks. **d,** E18.5 skin sections ($n = 3$ wild type and $n = 3$ $Irf6^{-/-}$) stained with Nile red fluorescent dye, which indicates polar lipids in red and non-polar lipids in green. 100 × 100 μm. **e,** Profiling of 12 classes of lipids in wild-type versus $Irf6^{-/-}$ E16.5 skin ($n = 4$ each). The $x$ axis

denotes the sum(log2(abundance in $Irf6^{-/-}$/abundance in wild type)) per analyte; values less than 0 indicate a decrease, and values greater than 0 an increase, in $Irf6^{-/-}$ versus wild-type skin. Green bars denote a significant false discovery rate (FDR). Moderated $t$-test (two-sided) with Benjamini–Hochberg correction. CE, cholesteryl esters; DAG, diacylglycerols; HCER, hexosylceramides; LPC, lysophosphatidylcholines; LPE, lysophosphatidylethanolamines; PC, phosphatidylcholines; PE, phosphatidylethanolamines; SM, sphingomyelins.; FC, fold change. **f, g,** Volcano plots showing CER (**f**) or TAG (**g**) species in wild-type versus $Irf6^{-/-}$ E16.5 skin ($n = 4$ each). The $x$ axis denotes log2(analyte abundance in $Irf6^{-/-}$/analyte abundance in wild type), and the $y$ axis indicates the FDR. CERs or TAGs that were significantly decreased in $Irf6^{-/-}$ versus wild-type skin are red. Moderated $t$-test (two-sided) with Benjamini–Hochberg correction.

corneum lipids[29,30] (*Pnpla1*, *Cers3*, *Sptlc3* and *Rora*) (Extended Data Figs. 8g, 9a). Underscoring the importance of lipid synthesis in the function of the skin barrier, mutation of *PNPLA1* or *CERS3* in humans results in barrier disruption and skin defects[31,32]. Accordingly, mice that are deficient in these genes have a paucity of epidermal lipids, lack a permeability barrier and die soon after birth[31–33].

Next, we determined whether lipid defects contributed to the barrier dysfunction that was observed in $Irf6^{-/-}$ embryos. Staining with Nile red revealed a marked absence of non-polar lipids in the outer layer of the skin of $Irf6^{-/-}$ embryos (Fig. 4d). Quantification of twelve classes of lipids in the skin of E16.5 embryos showed that six of these were significantly decreased in $Irf6^{-/-}$ compared to wild-type skin (Fig. 4e, Extended Data Fig. 9b), including ceramides (CERs) and their precursors, triacylglycerols (TAGs) (Fig. 4f, g). Ceramides were notable not only because they are required for a functional permeability barrier, but also because the IRF6 target genes *Pnpla1*, *Cers3* and *Sptlc3* encode enzymes that are involved in various stages of ceramide biosynthesis[29,31]. *Cers3* is specifically required for the synthesis of sphingolipids with ultra-long-chain acyl moieties (that is, with a carbon chain that has 26 or more carbon atoms)[33], and accordingly, $Irf6^{-/-}$ skin lacked sphingolipid species with a carbon-chain length of 26 or greater (Extended Data Fig. 9c, d). As predicted given the role of RIPK4 upstream of IRF6, analysis of the skin of $Ripk4^{D161N/D161N}$ E16.5 embryos also revealed an absence of non-polar lipid lamellar membranes and a significant decrease in three classes of lipids, including CERs, dihydroceramides (DCERs) and lactosylceramides (LCERs) (Extended Data Fig. 10a–c).

Thus, IRF6 drives a transcriptional program that is essential for the lipid composition of the stratum corneum and for a functional epidermal barrier. Our findings provide a mechanistic explanation for why genetic defects in the RIPK4–IRF6 axis, including IRF6-responsive transcripts, lead to a disrupted epidermal barrier and life-threatening ectodermal defects.

## Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-019-1615-3.

1. Fuchs, E. Skin stem cells: rising to the surface. *J. Cell Biol.* **180**, 273–284 (2008).
2. Holland, P. et al. RIP4 is an ankyrin repeat-containing kinase essential for keratinocyte differentiation. *Curr. Biol.* **12**, 1424–1428 (2002).
3. Ingraham, C. R. et al. Abnormal skin, limb and craniofacial morphogenesis in mice deficient for interferon regulatory factor 6 (Irf6). *Nat. Genet.* **38**, 1335–1340 (2006).
4. Kondo, S. et al. Mutations in *IRF6* cause Van der Woude and popliteal pterygium syndromes. *Nat. Genet.* **32**, 285–289 (2002).
5. Mitchell, K. et al. Exome sequence identifies *RIPK4* as the Bartsocas–Papas syndrome locus. *Am. J. Hum. Genet.* **90**, 69–75 (2012).
6. Rountree, R. B. et al. RIP4 regulates epidermal differentiation and cutaneous inflammation. *J. Invest. Dermatol.* **130**, 102–112 (2010).

7. De Groote, P. et al. A novel RIPK4–IRF6 connection is required to prevent epithelial fusions characteristic for popliteal pterygium syndromes. *Cell Death Differ.* **22**, 1012–1024 (2015).
8. Kalay, E. et al. Mutations in *RIPK4* cause the autosomal-recessive form of popliteal pterygium syndrome. *Am. J. Hum. Genet.* **90**, 76–85 (2012).
9. Massoud, A. A., Ammaari, A. N., Khan, A. S. S., ven Katraman, B. & Teebi, A. S. Bartsocas–Papas syndrome in an Arab family with four affected sibs: further characterization. *Am. J. Med. Genet.* **79**, 16–21 (1998).
10. Van Der Woude, A. Fistula labii inferioris congenita and its association with cleft lip and palate. *Am. J. Hum. Genet.* **6**, 244–256 (1954).
11. Butali, A. et al. Novel *IRF6* mutations in families with Van Der Woude syndrome and popliteal pterygium syndrome from sub-Saharan Africa. *Mol. Genet. Genomic Med.* **2**, 254–260 (2014).
12. Gorlin, R. J., Sedano, H. O. & Cervenka, J. Popliteal pterygium syndrome. a syndrome comprising cleft lip-palate, popliteal and intercrural pterygia, digital and genital anomalies. *Pediatrics* **41**, 503–509 (1968).
13. Richardson, R. J. et al. Irf6 is a key determinant of the keratinocyte proliferation-differentiation switch. *Nat. Genet.* **38**, 1329–1334 (2006).
14. Kwa, M. Q. et al. Receptor-interacting protein kinase 4 and interferon regulatory factor 6 function as a signaling axis to regulate keratinocyte differentiation. *J. Biol. Chem.* **289**, 31077–31087 (2014).
15. Huang, C. S. et al. Crystal structure of Ripk4 reveals dimerization-dependent kinase activity. *Structure* **26**, 767–777 (2018).
16. Dassule, H. R., Lewis, P., Bei, M., Maas, R. & McMahon, A. P. Sonic hedgehog regulates growth and morphogenesis of the tooth. *Development* **127**, 4775–4785 (2000).
17. Kousa, Y. A., Moussa, D. & Schutte, B. C. IRF6 expression in basal epithelium partially rescues *Irf6* knockout mice. *Dev. Dyn.* **246**, 670–681 (2017).
18. Urwyler-Rösselet, C. et al. Keratinocyte-specific ablation of RIPK4 allows epidermal cornification but impairs skin barrier formation. *J. Invest. Dermatol.* **138**, 1268–1278 (2018).
19. Chen, W. et al. Insights into interferon regulatory factor activation from the crystal structure of dimeric IRF5. *Nat. Struct. Mol. Biol.* **15**, 1213–1220 (2008).
20. Matsuzawa, N. et al. Two missense mutations of the *IRF6* gene in two Japanese families with popliteal pterygium syndrome. *Am. J. Med. Genet.* **152A**, 2262–2267 (2010).
21. Botti, E. et al. Developmental factor IRF6 exhibits tumor suppressor activity in squamous cell carcinomas. *Proc. Natl Acad. Sci. USA* **108**, 13710–13715 (2011).
22. Liu, H. et al. Irf6 directly regulates Klf17 in zebrafish periderm and Klf4 in sign oral epithelium, and dominant-negative KLF4 variants are present in patients with cleft lip and palate. *Hum. Mol. Genet.* **25**, 766–776 (2016).
23. Little, H. J. et al. Missense mutations that cause Van der Woude syndrome and popliteal pterygium syndrome affect the DNA-binding and transcriptional activation functions of IRF6. *Hum. Mol. Genet.* **18**, 535–545 (2009).
24. Voigt, P., Tee, W. W. & Reinberg, D. A double take on bivalent promoters. *Genes Dev.* **27**, 1318–1338 (2013).
25. de la Garza, G. et al. Interferon regulatory factor 6 promotes differentiation of the periderm by activating expression of Grainyhead-like 3. *J. Invest. Dermatol.* **133**, 68–77 (2013).
26. Peyrard-Janvid, M. et al. Dominant mutations in *GRHL3* cause Van der Woude syndrome and disrupt oral periderm development. *Am. J. Hum. Genet.* **94**, 23–32 (2014).
27. Ting, S. B. et al. A homolog of *Drosophila* grainy head is essential for epidermal integrity in mice. *Science* **308**, 411–413 (2005).
28. Furuse, M. et al. Claudin-based tight junctions are crucial for the mammalian epidermal barrier: a lesson from claudin-1-deficient mice. *J. Cell Biol.* **156**, 1099–1111 (2002).
29. Li, S. et al. Transcription factor CTIP1/ BCL11A regulates epidermal differentiation and lipid metabolism during skin development. *Sci. Rep.* **7**, 13427 (2017).
30. Dai, J., Brooks, Y., Lefort, K., Getsios, S. & Dotto, G. P. The retinoid-related orphan receptor RORα promotes keratinocyte differentiation via FOXN1. *PLoS ONE* **8**, e70392 (2013).
31. Grond, S. et al. PNPLA1 deficiency in mice and humans leads to a defect in the synthesis of omega-O-acylceramides. *J. Invest. Dermatol.* **137**, 394–402 (2017).
32. Eckl, K. M. et al. Impaired epidermal ceramide synthesis causes autosomal recessive congenital ichthyosis and reveals the importance of ceramide acyl chain length. *J. Invest. Dermatol.* **133**, 2202–2211 (2013).
33. Jennemann, R. et al. Loss of ceramide synthase 3 causes lethal skin barrier disruption. *Hum. Mol. Genet.* **21**, 586–608 (2012).

## METHODS

**Mice.** All mouse studies complied with relevant ethical regulations and were approved by the Genentech institutional animal care and use committee. All mouse alleles were maintained on a C57BL/6N genetic background.

$Ripk4^{+/D161N}$ mice were derived from a FLEX model generated by Genoway (France) using C57BL/6 ES cells. In brief, GAC encoding D161 was changed to AAC in exon 3. Remnant $lox511$ and $loxP$ recombination sites are retained 5′ and 3′ of exon 3, respectively. RIPK4 conditional knockout mice ($Ripk4^{+/loxp}$) were generated at Genentech using C57BL/6 C2 embryonic stem (ES) cells. A $loxP$ site was placed 362 bp upstream of exon 2. An FRT-flanked PGK promoter–Neo cassette and second $loxP$ was placed 203 bp downstream of exon 2. The Neo cassette was excised in ES cells using Flp recombinase.

IRF6 conditional knockout mice ($Irf6^{+/loxp}$) were generated at Genentech using C2 ES cells. A $loxP$ site was placed 151 bp upstream of exon 3. An FRT-flanked PGK promoter–Neo cassette and second $loxP$ were placed 157 bp downstream of exon 3. The Neo cassette was excised in ES cells using Flp. $Irf6^{+/-}$ mice were derived from $Irf6^{+/loxp}$ mice by excision of exon 3 using HTN-Cre treatment of one-cell embryos during in vitro fertilization[34].

$Irf6^{+/S413A,S424A}$ mice were generated at Genentech using C2 ES cells. TCC was changed to GCC for both point mutations in exon 8. An FRT-flanked Neo cassette placed upstream of exon 8 was excised in ES cells using Flp. $Irf6^{+/S413E,S424E}$ mice were generated at Genentech using C2 ES cells. TCC was changed to GAA for both point mutations in exon 8. An FRT-flanked Neo cassette placed upstream of exon 8 was excised in ES cells using Flp.

$Ripk4^{D161N}$ genotyping primers (5′-GTCTGGAGTGCAGCCCTTCTGT TTGG-3′, 5′-TTCTGGAAACTGCTGCTCAGGGTAGGGAG-3′) amplified 251-bp wild-type and 363-bp $Ripk4^{D161N}$ fragments. $Ripk4^{loxp}$ genotyping primers (5′-TGGGGATGATGTCAGTACCC-3′, 5′-GAGTGAGCAACCAGGATGCT-3′ and 5′-CACCCTGAAGCAGAGCAGGA-3′) amplified 274-bp wild-type, 217-bp $Ripk4^-$ and 360-bp $Ripk4^{loxp}$ fragments. $Irf6^{loxp}$ genotyping primers (5′-AATGTCGCCCCAGACTTCAGCTTCAG-3′, 5′-GTGTGAATGCTGACCACTATGGAGAAC-3′ and 5′-GAACACAATAGCTT GCACGGGTCATGTC-3′) amplified 251-bp wild-type, 363-bp $Irf6^-$ and 285-bp $Irf6^{loxp}$ fragments. $Irf6^{S413A,S424A}$ genotyping primers (5′-CTAAGTAAGCCACCACCTA-3′ and 5′- AAGAGCACACAAGTTTCC-3′) amplified 236-bp wild-type and 270-bp $Irf6^{S413A,S424A}$ fragments. $Irf6^{S413E,S424E}$ genotyping primers (5′- CATCCAGGTTCCTTTCTTTGA-3′ and 5′-ACACGC ACAGCAGTCT-3′) amplified 448-bp wild-type and 482-bp $Irf6^{S413E,S424E}$ fragments.

The K14-Cre mouse strain has been described previously[16]. The B6N. Cg-Tg(KRT14-cre)1Amc/J mouse line was imported from JAX, with permission from Harvard University, and crossed to the RIPK4 conditional knockout and IRF6 conditional knockout strains to generate $Ripk4^{EKO}$ and $Irf6^{EKO}$ mice, respectively. The K14-Cre allele was always kept in the heterozygous state, and was always paternally inherited, as the human K14 promoter is transcriptionally active in mouse oocytes and the enzyme remains active until after fertilization[35].

For timed pregnancies, males and females from 6 to 26 weeks of age were set up and E15.5–E18.5 embryos and P0 pups were analysed (both males and females). Embryos were considered to be at E0.5 the morning that a vaginal plug was detected.

No statistical methods were used to predetermine sample size. The experiments were not randomized and the investigators were not blinded to allocation during experiments and outcome assessment.

**Statistics.** Unless stated otherwise, statistical analysis of the results was performed using an unpaired, two-tailed Fisher's exact test. A 95% confidence interval was used for statistics and $P < 0.05$ was considered significant. All statistical analyses were performed using GraphPad Prism 6.

**Outside-in barrier assays.** The toluidine blue penetration assay was carried out as previously reported[36]. In brief, untreated and unfixed embryos were passaged through a chilled methanol gradient, rinsed in water and then immersed in 0.1% toluidine blue solution in water for 2 min.

**Inside-out barrier assays.** E18.5 embryos were removed from pregnant dams, resuscitated and placed in a humidified chamber (55–70% humidity), pre-warmed to 30 °C. Neonates were then weighed at 15-min intervals, for a total of 150 min. This protocol was based on a previous study[37].

**Histology, IHC and immunofluorescence.** Formalin-fixed paraffin-embedded tissue sections were labelled with rabbit anti-cytokeratin-14 (PRB-155P, Biolegend, 0.1 µg ml$^{-1}$), rabbit anti-cytokeratin-10 (PRB-159P, Biolegend, 1 µg ml$^{-1}$), mouse anti-Ki67 (clone SP6, RM-9106-S, Thermo Fisher, 1:200) or rabbit anti-IRF6 (GEN168NP-D5, Genentech, 0.5 µg ml$^{-1}$). IRF6 staining used target retrieval solution, pH 9 (S2368, Agilent) for antigen retrieval, and the other antibodies used the citrate-based target retrieval solution, pH 6 (S1699, diluted 1:10 in H$_2$O, Agilent). ABC Peroxidase Elite (Vector Laboratories) detection with DAB chromogen was used for all IHC.

For Nile red staining, unfixed frozen sections of fetal mouse skin were stained as described previously[38]. Two-colour laser scanning confocal microscopy was used to visualize the lipid lamellae in the epidermis.

For all histology, IHC and immunofluorescence experiments, at least three mice per genotype were analysed.

**Mammalian cell culture.** Human 293T cells (Genentech, 129641, 293T GNE) were cultured in high-glucose DMEM supplemented with 2 mM glutamine, 50 U ml$^{-1}$ penicillin, 50 µg ml$^{-1}$ streptomycin, 10% heat-inactivated fetal calf serum, 50 µM 2-mercaptoethanol, 1× non-essential amino acids solution (Thermo Fisher) and 10 mM HEPES pH 7.2, at 37 °C and 5% CO$_2$.

For SILAC, 293T cells were heavy labelled by culturing in SILAC DMEM (Thermo Fisher) supplemented with 50 µg ml$^{-1}$ lysine, 40 µg ml$^{-1}$ arginine, 200 µg ml$^{-1}$ proline, 2 mM glutamine, 50 U ml$^{-1}$ penicillin, 50 µg ml$^{-1}$ streptomycin, 10% dialysed fetal bovine serum, 50 µM 2-mercaptoethanol, 1× non-essential amino acids solution (Thermo Fisher) and 10 mM HEPES pH 7.2, at 37 °C and 5% CO$_2$ for at least 9 days. The medium was changed every day.

For transient overexpression of proteins in 293T cells, DNA encoding full-length mouse RIPK4 or IRF6 was placed into a pRK mammalian expression vector using gene synthesis. RIPK4 was tagged with a C-terminal Myc tag. IRF6 was either untagged (for constructs used in luciferase assays), or contained a 3×Flag tag at its C terminus (for constructs used for mass spectrometry). Various mutants of these proteins were generated using the QuikChange Site-Directed Mutagenesis Kit (Agilent) and were sequence-verified.

293T cells were transfected using 10 µg of plasmid DNA and lipofectamine-2000 (Thermo Fisher), according to the manufacturer's instructions, for 48 h. Cells were collected, washed with ice-cold PBS and lysed using 10 mM HEPES-KOH pH 7.9, 10 mM KCl, 1.5 mM MgCl$_2$ and 1× halt protease and phosphatase inhibitor cocktail (Thermo Fisher) for 20 min on ice. 25 µl of 10% NP-40 was added to the lysate and it was centrifuged at 20,000$g$ for 1 min at 4 °C. The supernatant (cytosolic fraction) was set aside. The pellet was then treated with 420 mM NaCl, 20 mM HEPES-KOH pH 7.9, 1.5 mM MgCl$_2$, 0.2 mM EDTA, 25% glycerol and 1× halt protease and phosphatase inhibitor cocktail (Thermo Fisher) for 20 min on ice with frequent agitation, and further centrifugation was performed at 20,000$g$ for 5 min at 4 °C to extract the chromatin-bound fraction. This was pooled together with the cytosolic fraction, which was then used for immunoprecipitation followed by mass spectrometry.

293T stocks were tested for mycoplasma before and after cryopreservation. The cell line was authenticated using short tandem repeat (STR) profiles (Promega PowerPlex 16 System) and compared to external STR profiles of cell lines to determine cell-line ancestry.

**Western blotting.** Mouse-tissue lysates were prepared by mechanical homogenization of tissue using the Omni Bead Ruptor system (Omni International), according to the manufacturer's instructions, and the following lysis buffer: 1% Triton X-100, 10% glycerol, 135 mM NaCl, 1.5 mM MgCl$_2$, 1 mM EGTA, 20 mM Tris-HCl pH 7.5 and 1× halt protease and phosphatase inhibitor cocktail (Thermo Fisher). 293T cells were lysed in the same buffer. After homogenization, lysates were clarified by centrifugation at 20,000$g$ for 10 min at 4 °C and denatured by the addition of 1× NuPAGE LDS sample buffer (Thermo Fisher) and 10 mM DTT and incubation at 70 °C for 10 min.

Lysates were separated on a NuPAGE 4–12% Bis-Tris protein gel (Thermo Fisher) and immunoblotted using the following antibodies: anti-FLAG (Sigma, 1:5,000), anti-MYC (GeneTex, 1:2,000), anti-IRF6 (rabbit monoclonal antibody (mAb); GEN168NP-F1, Genentech, 1 µg ml$^{-1}$), anti-RIPK4 (rat mAb; 3E9.3.1; Genentech, 1 µg ml$^{-1}$) and anti-β-actin (CST, 13E5, 1:2,000).

All western blots presented in this paper are representative of three independent experiments.

**Immunoprecipitation.** Immunoprecipitation was carried out on transfected 293T cell lysates using anti-Flag-M2 magnetic beads (Sigma-Aldrich), according to the manufacturer's instructions. Immunoprecipitated protein was eluted by incubating the magnetic beads with 100 µl of 0.2 mg ml$^{-1}$ 3×Flag peptide (Sigma-Aldrich) at 30 °C for 1 h. The eluate was concentrated using an 0.5 ml Amicon Ultra 30K concentrator (Millipore) before being used for mass spectrometry.

**Luciferase assays.** 293T cells (100 µl) were seeded into each well of a flat-bottom 96-well plate at $2.5 \times 10^5$ cells ml$^{-1}$ the day before transfection. The next day, each well was transfected with the following amounts of plasmid DNA: 90 ng of pGL3 (containing an IFNβ promoter driving the expression of $luc+$, a modified firefly luciferase), 10 ng of pRL-TK (containing the wild-type $Renilla$ luciferase gene for normalization), 50 ng of a pRK vector encoding untagged mouse wild-type IRF6 (or various mutants generated using the QuikChange Site-Directed Mutagenesis Kit from Agilent) and 50 ng of a pRK vector encoding C-terminal Myc-tagged mouse RIPK4 (either wild-type or containing the D161N mutation), using lipofectamine-2000 (Thermo Fisher), according to the manufacturer's instructions, for 24 h. The same total amount of DNA was always included per well (200 ng), and if necessary an empty pRK vector was used to make up that amount.

Luciferase Assays were carried out using the Dual-Glo luciferase assay system (Promega) according to the manufacturer's instructions, and firefly and *Renilla* luminescence were read using an Envision plate reader. The ratio of firefly:*Renilla* luminescence was calculated for each well, and values were normalized to the reporter-only control. The experiment was always performed using technical triplicates, and final data are an average of at least four biological repeats.

**Quantitative reverse-transcription PCR.** RNA from mechanically disrupted embryonic skin tissue was isolated using a RNeasy mini-kit (Qiagen). Quantitative reverse-transcription PCR (RT–qPCR) was performed using One-step Real-time RT–PCR Mastermix (ABI) and the following Taqman probes: *Gapdh* (Mm99999915_g1), *Grhl3* (Mm01193339_m1), *Ocln* (Mm00500912_m1), *Pnpla1* (Mm01308771_m1), *Sptlc3* (Mm01278138_m1), *Cers3* (Mm03990709_m1) and *Rora* (Mm01173766_m1), all from Life Technologies. Values were normalized to *Gapdh* transcript levels, and gene expression was calculated relative to the wild type. Experiments were always performed as technical triplicates, and final data are an average of three biological repeats.

**Protein purification.** The mouse RIPK4 kinase domain (either wild type or the kinase-dead BPS mutant T184I) was purified as described previously[15].

Full-length mouse wild-type IRF6 (A2-Q467) was cloned into a modified pET vector (MilliporeSigma) with a tobacco etch virus (TEV)-cleavable N-terminal His8 tag and transformed into BL21 Star (DE3) *E. coli* (Thermo Fisher). Cells were grown in TB medium, induced with 0.4 mM isopropyl β-D-1-thiogalactopyranoside (IPTG; GoldBio) and expressed at 20 °C overnight, and the cells were collected by centrifugation. Following cell lysis, the soluble protein was purified over Ni-NTA agarose (Qiagen), followed by size-exclusion chromatography on a Superdex 75 16/60 column (GE Healthcare). Protein-containing fractions were analysed by SDS–PAGE and electrospray mass spectrometry to verify molecular weight. The N-terminal His8 tag was not removed before subsequent experiments.

**In vitro kinase assays.** Recombinant full-length mouse IRF6 (5 μM) and recombinant mouse RIPK4 kinase domain (either wild type or T184I; 0.5 μM) were incubated in a basic reaction buffer containing 50 mM Tris pH 7.5, 20 nM MgCl$_2$ and 100 μM ATP at 37 °C for either 5 or 30 min. The reaction was quenched by adding 1× NuPAGE LDS Sample Buffer (containing 10 mM DTT) and incubating at 70 °C for 10 min. Samples were then alkylated by treating with 20 mM iodoacetamide at room temperature in the dark for 20 min, loaded onto a NuPAGE 4–12% Bis-Tris protein gel, stained with SimplyBlue stain (Invitrogen) and destained in water, before being used for mass spectrometry.

**Lipidomics.** Skin tissue from wild-type and *Irf6*$^{-/-}$ E16.5 embryos ($n = 4$ of each) and wild-type and *Ripk4*$^{D161N/D161N}$ E16.5 embryos ($n = 4$ of each), weighing approximately 100 mg, were weighed and homogenized in 600 μl dichloromethane (DCM):methanol (1:1, v-v) using ceramic beads on a Omni Bead Ruptor 24 Homogenizer for 90 s. After centrifugation, 300 μl of the supernatant was transferred into a conical bottom glass tube and 1 ml H$_2$O, 0.75 ml DCM and 1.85 ml methanol were added to the supernatant to form a single phase. After a 30-min incubation, isotope-labelled internal standards (SCIEX Lipidyzer Platform kits, 5040156) were added to the mixture, followed by 0.9 ml DCM and 1 ml water. The mixture was centrifuged at 1,000g for 20 min to achieve phase separation. The bottom layer was collected into a clean glass tube, and the upper layer was re-extracted by adding 1.8 ml of DCM. The bottom layer of the second extraction was combined with the first and dried under a gentle stream of nitrogen. The residue was reconstituted in 300 μl DCM:methanol (1:1) containing 10 mM ammonium acetate for direct infusion and analysis on a SelexION enabled 6500 QTRAP (Sciex), using Lypidyzer platform methods (Sciex).

Before statistical analysis, samples were normalized by their wet weight to correct for variability in preparation. The data were then analysed using a linear model implemented in the Limma package[39] for the R programming language[40]. Differences between wild-type and *Irf6*$^{-/-}$ samples were tested using a moderated version of the *t*-test. Resulting *P* values were adjusted for multiple comparisons by controlling the FDR using the Benjamini–Hochberg method. Data were uploaded to a Spotfire (TIBCO Software) dashboard for interrogation and plot generation.

**RNA-seq.** RNA-seq profiling was performed to determine whether IRF6 and RIPK4 perform a similar function in keratinocyte differentiation. For this we used skin samples from wild-type, *Irf6*$^{-/-}$, *Irf6*$^{S413A,S424A/S413A,S424A}$ and *Ripk4*$^{D161N/D161N}$ E15.5 mouse embryos. Wild-type ($n = 3$) and *Irf6*$^{-/-}$ ($n = 5$) samples, as well as wild-type ($n = 3$) and *Ripk4*$^{D161N/D161N}$ ($n = 5$) samples, were litter matched. Wild-type ($n = 3$) and *Irf6*$^{S413A,S424A/S413A,S424A}$ ($n = 5$) were not litter matched. We also used skin samples from wild-type and *Irf6*$^{-/-}$ E16.5 embryos ($n = 3$ of each genotype; not litter matched).

RNA from mechanically disrupted skin tissue was isolated using the RNeasy mini-kit (Qiagen). The concentration of RNA samples was determined using NanoDrop 8000 (Thermo Fisher Scientific) and the integrity of RNA was determined by Fragment Analyzer (Advanced Analytical Technologies). Exactly 0.1 μg of total RNA was used as input material for library preparation using a TruSeq Stranded Total RNA Library Prep Kit (Illumina). The size of the libraries was

confirmed using 4200 TapeStation and High Sensitivity D1K screen tape (Agilent Technologies) and their concentration was determined by a qPCR-based method, using a library quantification kit (KAPA). The libraries were multiplexed and then sequenced on Illumina HiSeq4000 (Illumina) to generate 30 million single-end 50-base-pair reads.

Sequencing reads were aligned to the mouse genome (GRCm38/mm10) with gene models (Gencode M15) using GNAP (v.2013-11-10; research-pub.gene.com/gmap/) and the following parameters: -M 2 -n 10 -B 2 -i 1 -N 1 w 200,000 -E 1–pairmax-rna = 200,000–clip-overlap. Gene-expression levels were computed by summing the number of reads that mapped to exons of Gencode genes. Differential expression between groups of samples was computed using the R limma package[41] after transformation from raw counts using the voom function. For visualization, counts were normalized using size factors (as described by DESeq2[42]) to account for library size and gene length. For visualization purposes, reads per kilobase of transcript per million mapped reads (RPKM) values were calculated for each gene in the Gencode genes set. The R package EGSEA[43] was used to perform gene enrichment analysis.

Genes were considered differentially expressed if the log$_2$(expression in mutant/expression in wild-type) was greater than 1 or less than −1 and the adjusted P value less than 0.05. The data are available at the Gene Expression Omnibus (GEO) under accession code GSE124067.

**ChIP–seq.** Snap-frozen skin samples from wild-type and *Irf6*$^{-/-}$ E16.5 mouse embryos (200 mg, $n = 2$ of each genotype) were sent to Active Motif for ChIP–seq. ChIP was performed with validated antibodies against H3K27me3 (Active Motif 39155), H3K27ac (Active Motif 39133) or H3K4me3 (Active Motif 39159). Illumina sequencing libraries were prepared from the ChIP and input DNA using the standard consecutive enzymatic steps of end polishing, dA addition and adaptor ligation using Active Motif's custom liquid handling robotics pipeline. After the final 15-cycle PCR amplification step, the resulting DNA libraries were quantified and sequenced on NextSeq 500.

Sequencing reads were aligned to the mouse genome (GRCm38/mm10) using GSNAP (v.2013-11-10) and the following parameters: -M 2 -n 10 -B 2 -i 1–pairmax-dna = 1,000–terminal-threshold = 1,000–gmap-mode = none–clip-overlap. Mapped reads then were assessed for peaks relative to the input controls using Macs2 (v.2.1.0) callpeak function[44]. Peak-fold enrichment was calculated using Macs2, using a sliding window across the genome and assessing read counts relative to expected background. HOMER[45] (v.4.7) is used to identify the motif that are enriched in these peaks. The Gviz R package[46] was used to visualize sample coverage at *Grhl3*, *Occludin* and *Pnpla1* loci. The R package ComplexHeatmap[47] was used to plot the heat maps. Samples were normalized to sequencing depth.

**ATAC-seq.** Skin was collected from wild-type and *Irf6*$^{-/-}$ E16.5 mouse embryos (100 mg, $n = 2$ of each genotype) and placed in L-15 medium containing 1× DNase buffer and 200 μl ml$^{-1}$ DNase for 20 min at room temperature. Samples were then cryopreserved by placing them in L-15 medium containing 10% FBS and 5% DMSO and freezing them at a slow cooling rate. ATAC-seq was performed by Active Motif. In brief, the tissue was manually disassociated, isolated nuclei were quantified using a haemocytometer and 100,000 nuclei were tagmented as previously described[48], with some modifications[49], using the enzyme and buffer provided in the Nextera Library Prep Kit (Illumina). Tagmented DNA was then purified using the MinElute PCR purification kit (Qiagen), amplified with 10 cycles of PCR and purified using Agencourt AMPure SPRI beads (Beckman Coulter). Resulting material was quantified using the KAPA Library Quantification Kit for Illumina platforms (KAPA Biosystems), and sequenced with PE42 sequencing on the NextSeq 500 sequencer (Illumina).

Raw paired-end FASTQ files were aligned to the mouse reference genome (GRCm38/mm10) using GNAP (-M 2 -n 10 -B 2 -i 1–pairmax-dna = 1,000–terminal-threshold = 1,000–gmap-mode = none–clip-overlap). Reads that mapped to the mitochondrial chromosome or to blacklisted regions were removed from the downstream analysis. The data was further processed using the standard ENCODE pipeline with minor modifications (https://www.encodeproject.org/atac-seq/#standards). Peaks were called using Macs2[44], first at sample level and then at group level by pooling samples together. Peaks from pooled samples that were also called in the biological replicates were retained for downstream analysis.

**Identifying putative IRF6 targets.** HOMER[45] was run with default parameters on peaks from H3K4me3, H3K27me3 and H3K27ac in *Irf6*$^{-/-}$ and wild-type backgrounds. Only the H3K4me3 data set (both *Irf6*$^{-/-}$ and wild type) yielded an ISRE motif. FIMO[50] was used to scan for the location of the ISRE motif genome-wide. Motifs that intersected with ATAC-seq peaks from *Irf6*$^{-/-}$ or wild-type datasets were used for downstream analysis. We assigned motifs to the nearest gene if a motif was within 1,000 bp upstream of the gene start to 1,000 bp downstream of the gene end. This gave us 869 putative IRF6 targets. To further characterize a more confident set of IRF6 targets, we intersected putative IRF6 targets with genes that were differentially expressed in the RNA-seq data comparing *Irf6*$^{-/-}$ to wild type

($\log_2$(expression in $Irf6^{-/-}$/expression in wild type) $> 1$ or $< -1$) and adjusted $P < 0.05$. This gave us a list of 66 high-confidence IRF6 targets.

**Clustering analysis.** Signal intensities for each replicate of H3K4me3 and H3K27me3 from wild-type samples were first calculated by summing up read counts in a 5,000-bp window around the gene start. The counts were then normalized by the total mapped reads per sample followed by $\log_2$ transformation. Genes with average signal less than 0.8 were filtered out. $k$-means clustering was then used to cluster the genes into four groups.

**Mass spectrometry.** Samples were reduced with 10 mM DTT in $1\times$ NuPAGE LDS sample buffer at 95 °C for 10 min followed by alkylation with 20 mM iodo-acetamide at room temperature in the dark for 20 min. Samples were loaded onto a NuPAGE 4–12% Bis-Tris protein gel. Protein was stained with SimplyBlue stain (Invitrogen) and destained in water.

Trypsin and chymotrypsin digestion were used to identify and quantify IRF6 pS413 and pS424 and to identify new RIPK4-dependent phosphorylation sites on IRF6. Lys-C digestion was used to identify and quantify IRF6 pS90. The protein bands were excised and further de-stained in 50 mM ammonium bicarbonate ($NH_4HCO_3$)/30% acetonitrile (ACN) and dehydrated in 100% ACN. Gel pieces were rehydrated with 10 ng $\mu l^{-1}$ trypsin or chymotrypsin or Lys-C in 25 mM $NH_4HCO_3$ and chilled on ice for 1 h. Excess trypsin, chymotrypsin and Lys-C solutions were removed and digestions were performed in 25 mM $NH_4HCO_3$ at 37 °C overnight. Peptides were extracted with 0.1% trifluoroacetic acid (TFA)/ACN. For absolute quantification (AQUA) analysis, Heavy isotopically labelled peptides flanking pS413 & pS424 were spiked into the digest mixture such that 150 fmol was loaded onto the $C_{18}$ column (Supplementary Table 1). The peptide mixture was dried to completion and re-suspended in 0.1% formic acid for tandem mass spectrometric analysis.

Samples were injected using an auto-sampler for separation by reverse phase chromatography on a NanoAcquity UPLC system (Waters). Peptide digest was loaded onto a Symmetry $C_{18}$ column (1.7 mm BEH-130, 0.1 $\times$ 100 mm, Waters) with a flow rate of 1 $\mu l\ min^{-1}$ and a gradient of 0% to 25% Solvent B (where Solvent A is 0.1% FA/2% ACN/water and Solvent B is 0.1% formic acid/2% water/ACN) applied over 60 min with a total analysis time of 90 min. Peptides were eluted directly into an Advance CaptiveSpray ionization source (Michrom BioResources/Bruker) with a spray voltage of 1.3 kV and were analysed using an LTQ Orbitrap Elite mass spectrometer (Thermo Fisher). Precursor ions were analysed in the Orbitrap at 60,000 resolution; tandem mass spectrometry was performed in the LTQ with the instrument operated in data-dependent mode whereby the top 15 most-abundant ions were subjected to fragmentation.

Extracted ion chromatograms of the endogenous (light) and heavy labelled peptides were generated and the area under the curve (AUC) was integrated. Abundance measurements were calculated for each analyte peptide as the ratio of light-to-heavy peptides. For comprehensive mapping and quantification of IRF6 phosphorylation sites, spectra from tandem mass spectometry were searched using Mascot search algorithm (Matrix Science) against a concatenated target–decoy database consisting of mouse protein sequences, common contaminants and the reversed sequence of each protein. Precursor and fragment ion mass tolerance were set at 50 p.p.m. and 0.8 Da, respectively, with trypsin and chymotrypsin specificity and a maximum of three miscleavages. Variable modifications included methionine oxidation (+15.9949 Da), iodoacetamide adduct for cysteine residues (+57.0215 Da), phosphorylation (+79.9663 Da) for serine, threonine & tyrosine residues, heavy lysine (+8.0142 Da) and heavy arginine (+10.0083 Da). Peptide assignments were filtered at 1% FDR at peptide level using linear discriminant analysis. Phosphorylation sites were localized by Ascore algorithm[51]. All tryptic, chymotryptic and Lys-C peptides containing Lys and Arg residues were quantified using VistaQuant algorithm as previously described[52]. Quantified peptides were exported using a VQ confidence cut-off of 83 for statistical analysis (for experiments performed in two biological replicates) and data visualization using R (version 3.5.0). Quantification of phosphorylation sites was summarized by modelling all peptide features containing the respective site quantified in both biological replicates, regardless of digestion conditions. A $\log_2$-transformed ratio of mutant versus wild type and a $P$ value were reported for each site (or combination of sites, if multiple phosphorylation sites were observed from the same peptide feature) using a linear mixed effect model (R package nlme). For unphosphorylated peptides, heavy and light AUCs quantified from all biological replicates and digestion conditions were aggregated, respectively, in each peptide form. A ratio of mutant versus wild type was then calculated and $\log_2$-transformed. All results were plotted using the R package ggplot2.

**Reporting Summary.** Further information on research design is available in the Nature Research Reporting Summary linked to this article.

## Data availability

The data sets generated in this study are available from the corresponding authors on reasonable request. Gel source data can be found in Supplementary Fig. 1. Source data for all graphs are provided.

## Code availability

All custom code can be accessed at https://github.com/rreja/Oberbeck_Nature_2019.

34. Ryder, E. et al. Rapid conversion of EUCOMM/KOMP-CSD alleles in mouse embryos using a cell-permeable Cre recombinase. *Transgenic Res.* **23**, 177–185 (2014).
35. Hafner, M. et al. Keratin 14 Cre transgenic mice authenticate keratin 14 as an oocyte-expressed protein. *Genesis* **38**, 176–181 (2004).
36. DiTommaso, T. & Foijer, F. Barrier function assay. *Bio Protoc.* **4**, e1133 (2014).
37. Gareus, R. et al. Normal epidermal differentiation but impaired skin-barrier formation upon keratinocyte-restricted IKK1 ablation. *Nat. Cell Biol.* **9**, 461–469 (2007).
38. Hanley, K. et al. Glucocorticoid deficiency delays stratum corneum maturation in the fetal mouse. *J. Invest. Dermatol.* **111**, 440–444 (1998).
39. Smyth, G. K. in *Bioinformatics and Computational Biology Solutions Using R and Bioconductor* (eds Gentleman, R., Carey, V. J., Huber, W., Irizarry, R. A. & Dudoit, S.) 397–420 (Springer, New York, 2005).
40. R Core Team. *R: A language and environment for statistical computing.* (R Foundation for Statistical Computing, 2018).
41. Ritchie, M. E. et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **43**, e47 (2015).
42. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
43. Alhamdoosh, M. et al. Combining multiple tools outperforms individual methods in gene set enrichment analyses. *Bioinformatics* **33**, 414–424 (2017).
44. Zhang, Y. et al. Model-based analysis of ChIP–seq (MACS). *Genome Biol.* **9**, R137 (2008).
45. Heinz, S. et al. Simple combinations of lineage-determining transcription factors prime *cis*-regulatory elements required for macrophage and B cell identities. *Mol. Cell* **38**, 576–589 (2010).
46. Hahne, F. & Ivanek, R. Visualizing genomic data using Gviz and bioconductor. *Methods Mol. Biol.* **1418**, 335–351 (2016).
47. Gu, Z., Eils, R. & Schlesner, M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics* **32**, 2847–2849 (2016).
48. Buenrostro, J. D., Giresi, P. G., Zaba, L. C., Chang, H. Y. & Greenleaf, W. J. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat. Methods* **10**, 1213–1218 (2013).
49. Corces, M. R. et al. An improved ATAC-seq protocol reduces background and enables interrogation of frozen tissues. *Nat. Methods* **14**, 959–962 (2017).
50. Bailey, T. L. et al. MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res.* **37**, W202–W208 (2009).
51. Beausoleil, S. A., Villén, J., Gerber, S. A., Rush, J. & Gygi, S. P. A probability-based approach for high-throughput protein phosphorylation analysis and site localization. *Nat. Biotechnol.* **24**, 1285–1292 (2006).
52. Elias, J. E. & Gygi, S. P. Target-decoy search strategy for mass spectrometry-based proteomics. *Methods Mol. Biol.* **604**, 55–71 (2010).

**Author contributions** M.R.-G. and S.W. generated the $Ripk4^{+/loxp}$, $Irf6^{+/loxp}$, $Irf6^{+/S413A,S424A}$ and $Irf6^{+/S413E,S424E}$ mice; N.O. designed and performed experiments with help from A.M. and D.L.D; R.R. and Y.Z. performed bioinformatic analysis; C.S.H. purified recombinant proteins; V.C.P. performed mass spectrometry with help from K.Y., Q.L., A.B. and W.W; W.S. performed lipidomics; L.G.K. performed confocal microscopy; J.D.W. analysed histological data; and V.M.D. and K.N. helped with experimental design.
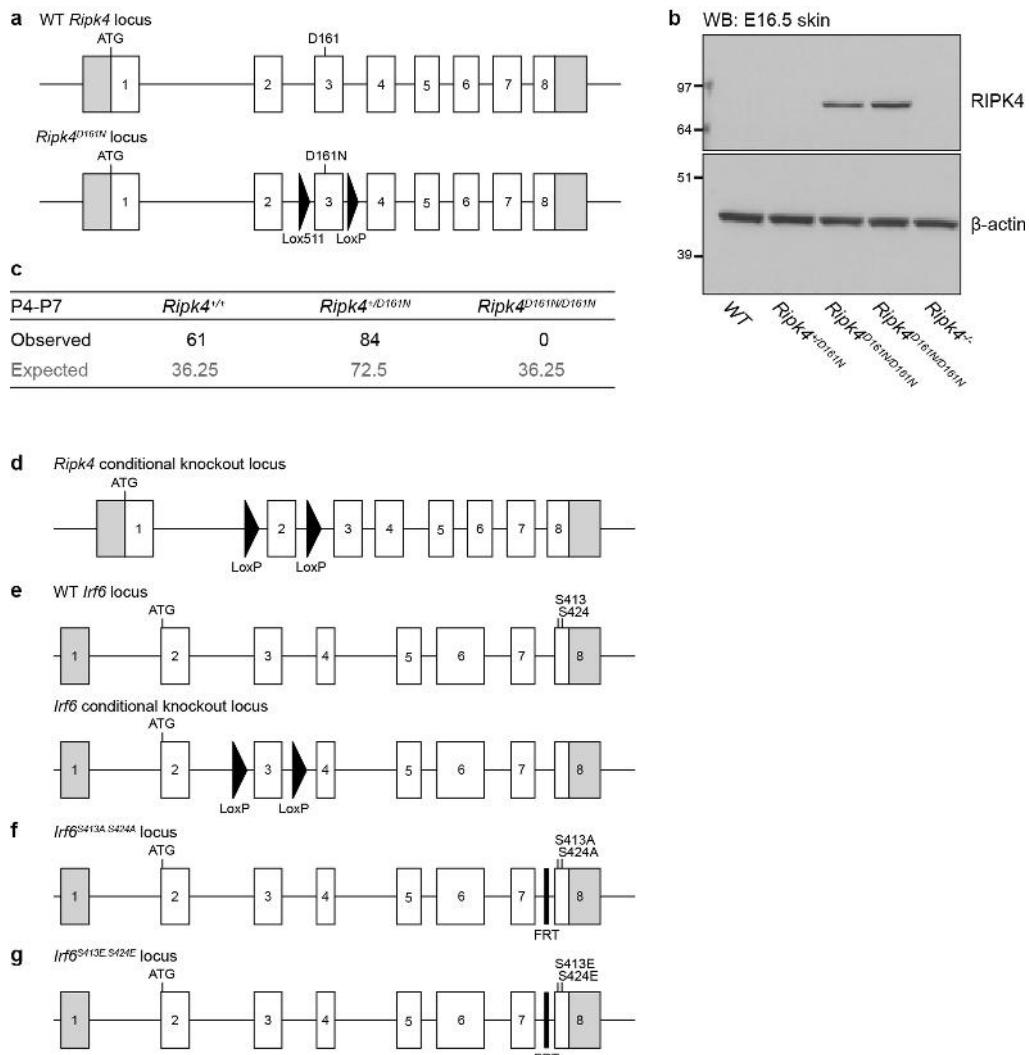
**Extended Data Fig. 1 | $Ripk4^{D161N/D161N}$ mice are neonatal lethal.**
**a**, Organization of the constitutive $Ripk4^{D161N}$ knock-in allele. Boxes represent exons. Untranslated regions are shaded grey. Not to scale. **b**, Western blots of wild-type, $Ripk4^{+/D161N}$ and $Ripk4^{D161N/D161N}$ E16.5 skin. Representative of three independent experiments. **c**, Table showing the observed and expected numbers at clipping (P4–P7) of offspring that were generated from intercrossing $Ripk4^{+/D161N}$ mice. **d–g**, Organization of the conditional $Ripk4$ knockout allele (**d**), conditional $Irf6$ knockout allele (**e**), constitutive $Irf6^{S413A,S424A}$ knock-in allele (**f**) and constitutive $Irf6^{S413E,S424E}$ knock-in allele (**g**). Boxes represent exons. Untranslated regions are shaded grey. Not to scale.
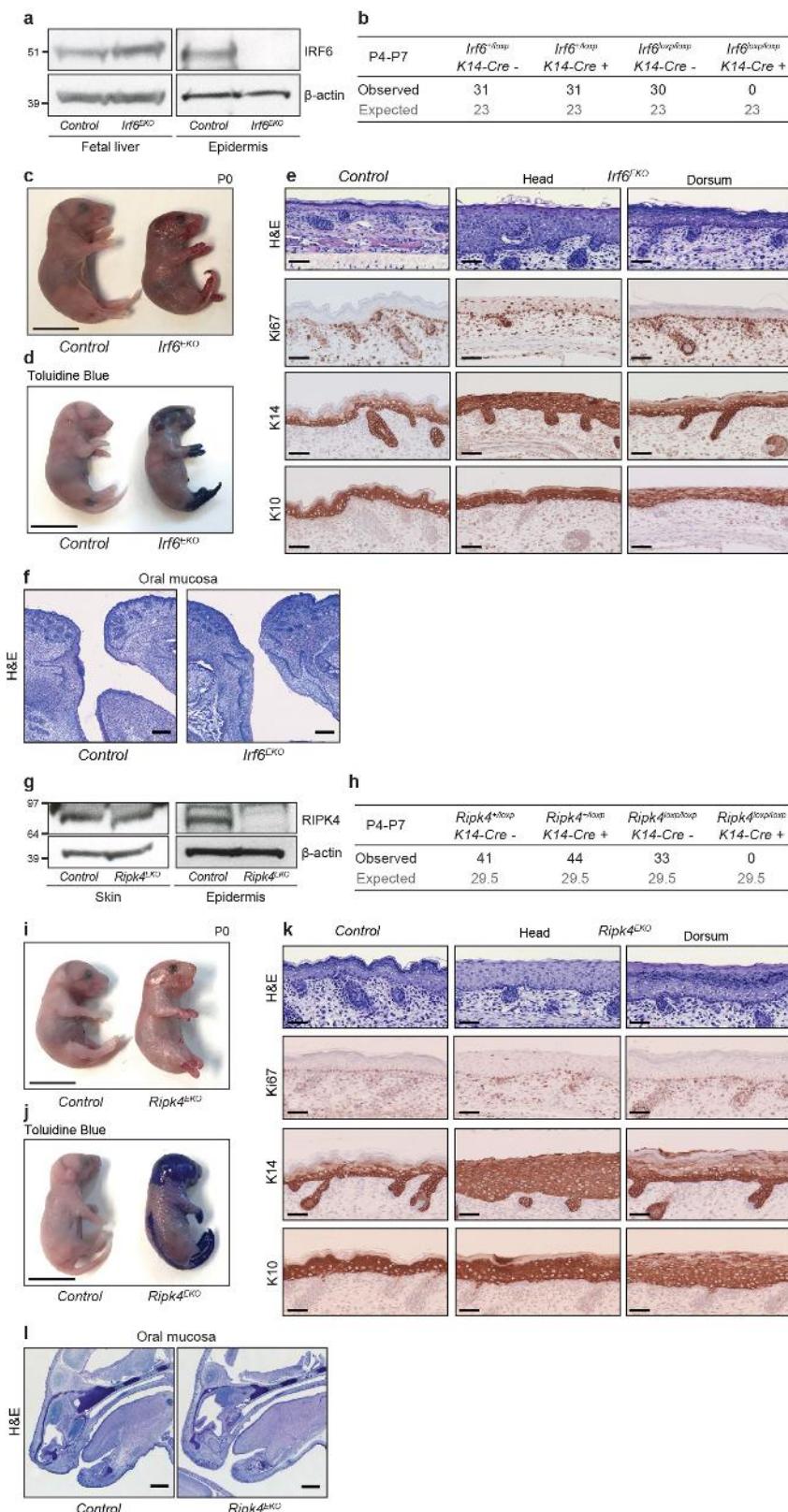
**Extended Data Fig. 2 |** See next page for caption.

**Extended Data Fig. 2 | Keratinocyte-specific deletion of *Irf6* or *Ripk4* demonstrates their cell-autonomous requirement for epidermal differentiation. a**, Western blots of epidermis and fetal liver from an E18.5 *Irf6*$^{EKO}$ embryo (K14-Cre$^+$; *Irf6*$^{loxp/loxp}$) and its K14-Cre$^+$; *Irf6*$^{+/loxp}$ littermate control. As expected, there is efficient deletion of IRF6 in the epidermis only. Representative of three independent experiments. **b**, Table showing the observed and expected numbers at clipping (P4–P7) of offspring that were generated from intercrossing *Irf6*$^{loxp/loxp}$; K14-Cre$^-$ females and *Irf6*$^{+/loxp}$; K14-Cre$^+$ males. No *Irf6*$^{EKO}$ mice were observed. **c**, P0 *Irf6*$^{EKO}$ and its littermate control ($n = 3$ each). Scale bar, 1 cm. **d**, Toluidine blue staining of P0 *Irf6*$^{EKO}$ and control embryos ($n = 3$ each). Scale bar, 1 cm. **e**, E18.5 skin sections from control and *Irf6*$^{EKO}$ embryos ($n = 3$ each) stained with H&E or antibodies against K10, K14 and Ki67. Scale bars, 50 μm. **f**, E18.5 sections from control and *Irf6*$^{EKO}$ embryos ($n = 3$ each) stained with H&E, showing no fusion of the squamous epithelium at the mouth. Scale bars, 200 μm. **g**, Western blots of epidermis and whole skin from an E18.5 *Ripk4*$^{EKO}$ embryo (K14-Cre$^+$; *Ripk4*$^{loxp/loxp}$) and its K14-Cre$^+$; *Ripk4*$^{+/loxp}$ littermate control. As expected, there is efficient deletion of *Ripk4* in the epidermis only. Representative of three independent experiments. **h**, Table showing the observed and expected numbers at clipping (P4–P7) of offspring that were generated from intercrossing *Ripk4*$^{loxp/loxp}$; K14-Cre$^-$ females and *Ripk4*$^{+/loxp}$; K14-Cre$^+$ males. No *Ripk4*$^{EKO}$ mice were observed. **i**, P0 *Ripk4*$^{EKO}$ and its littermate control ($n = 3$ each). Scale bar, 1 cm. **j**, Toluidine blue staining of P0 *Ripk4*$^{EKO}$ and control embryos ($n = 3$ each). Scale bar, 1 cm. **k**, E18.5 skin sections from control and *Ripk4*$^{EKO}$ embryos, stained with H&E or immunolabelled with antibodies against K10, K14 and Ki67. Scale bars, 50 μm. **l**, E18.5 sections from control and *Ripk4*$^{EKO}$ embryos stained with H&E, showing no fusion of the squamous epithelium at the mouth and no fusion of the tongue to the palate. Scale bars, 800 μm.

**Extended Data Fig. 3 | The epidermis of E18.5 $Ripk4^{D161N/D161N}Irf6^{-/-}$ embryos is indistinguishable from that of $Irf6^{-/-}$ embryos. a**, E18.5 sections with the indicated genotypes ($n = 3$ each) stained with H&E, showing fusion of the oral mucosa (arrows), irregular incisors with premature eruption (asterisks) and fusion of the stratified squamous portion of the stomach in the $Irf6^{-/-}$ and $Ripk4^{D161N/D161N}Irf6^{-/-}$ embryos. Scale bars, 200 µm (oral mucosa); 100 µm (stomach). **b**, E18.5 skin sections with the indicated genotypes ($n = 3$) immunolabelled with antibodies against K10, K14 and Ki67. Scale bars, 50 µm. **c**, Table showing the observed and expected numbers at clipping (P4–P7) of offspring that were generated from intercrossing $Irf6^{+/-}Ripk4^{+/D161N}$ and wild-type mice. $Irf6^{+/-}Ripk4^{+/D161N}$ offspring were weaned at the expected numbers.
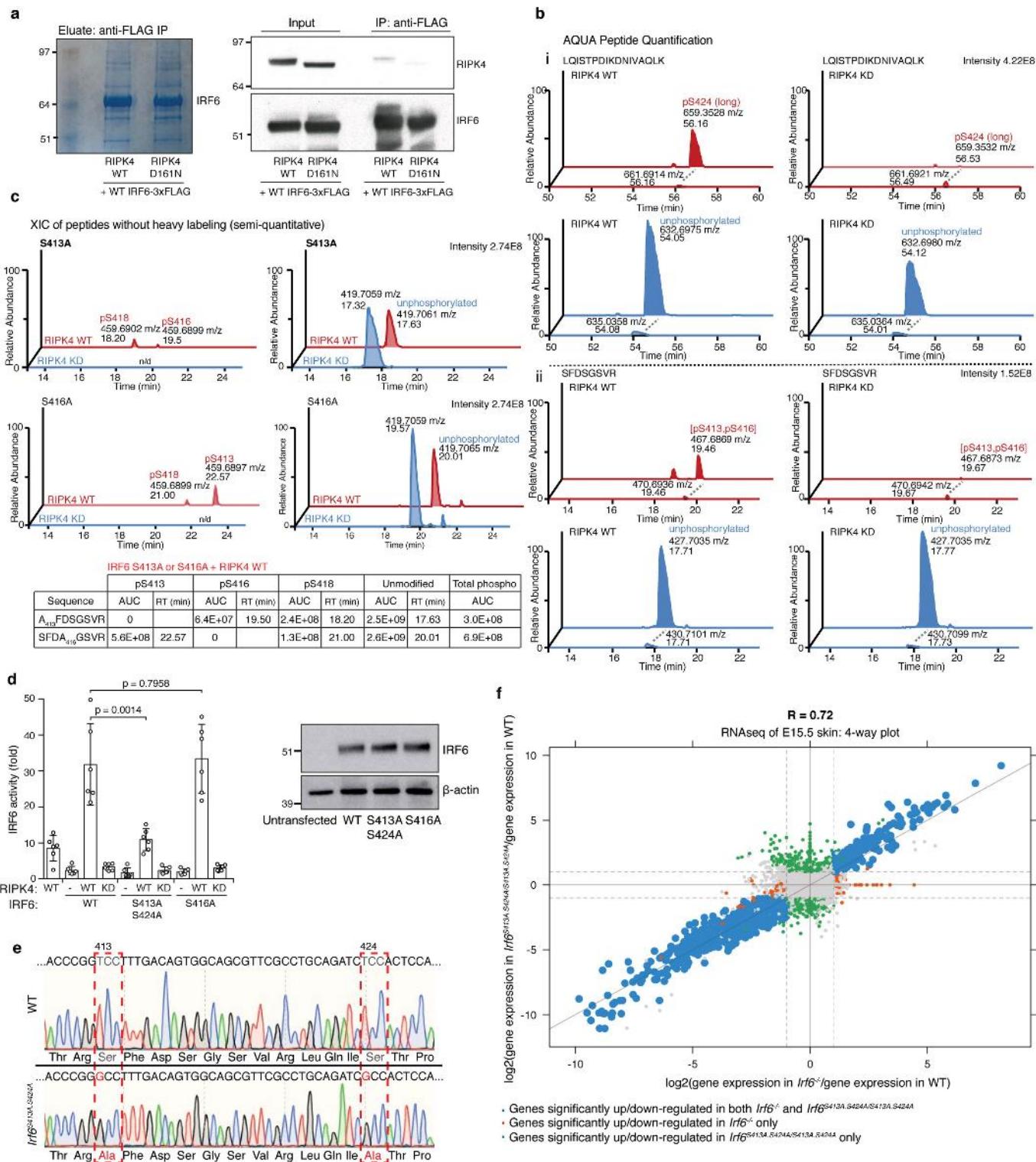
**Extended Data Fig. 4 |** See next page for caption.

**Extended Data Fig. 4 | Phosphorylation of IRF6 at Ser413 and S424 in vivo is essential for epidermal differentiation and development.** **a**, 3×FLAG-tagged IRF6 was affinity-purified from 293T cells co-expressing Myc-tagged RIPK4 or RIPK4(D161N) and then stained with SimplyBlue (left) or western blotted (right). IP, immunoprecipitation. Results representative of two independent experiments. **b**, Extracted ion chromatograms (XIC) of the phosphorylated peptides and their corresponding unmodified counterparts at pS424 (LQISTPDIKDNIVAQLK) (**i**) and pS(413,416) (SFDSGSVR) (**ii**); both the endogenous (light label) and the synthetic (heavy label) isotopically labelled peptides are shown. Plots are representative of two independent experiments. **c**, Absolute quantification at the pS413 and pS416 sites is difficult because peptides covering these sites co-elute. An experiment was therefore performed using IRF6(S413A) and IRF6(S416A) to distinguish between the levels of phosphorylation at these two sites. The resulting extracted ion chromatograms of the peptides without heavy labelling are shown, with the corresponding (AUC) values in the table below (RT, retention time). On the basis of the AUCs, there is more pS413 in the IRF6(S416A) sample than pS416 in the IRF6(S413A) sample. Given that the level of the unmodified form is similar, this suggests that S413 is the main phosphorylation site. This conclusion assumes that the ionization efficiency of these two peptides is the same, and is therefore semi-quantitative. **d**, Left, graph indicates activation of an IRF-responsive

luciferase reporter gene at 24 h after transfection of 293T cells with the indicated IRF6 and RIPK4 constructs. IRF6 activity is displayed as fold activity over reporter only. Data are mean ± s.d ($n = 6$). Unpaired, two-tailed Fisher's exact test with 95% confidence interval. Right, western blots show expression of the IRF6 phosphorylation mutants that were used in the luciferase assay. Representative of three independent experiments. **e**, Production of knock-in mutant mice expressing IRF6(S413A/S424A). Representative genomic sequencing of wild-type and homozygous ($Irf6^{S413A,S424A/\,S413A,S424A}$) mice. Nucleotides that encode Ser413 and Ser424 are highlighted by the dashed boxes, which indicate the wild-type GCC (Ser) and homozygous knock-in mutation TCC (Ala). **f**, Genome-wide four-way plot showing genes that have increased or decreased expression in $Irf6^{S413A,S424A/S413A,S424A}$ ($y$ axis, $n = 5$) or $Irf6^{-/-}$ ($x$ axis, $n = 5$) compared to the wild type ($n = 3$) in E15.5 skin. Each coloured dot represents a gene that met the cut-offs of an adjusted $P$ value <0.05 and a minimum twofold change in expression. Adjusted $P$ values were obtained using a moderated $t$-test (two-sided) with the Benjamini–Hochberg method for multiple comparisons. Genes that were altered significantly in expression in $Irf6^{-/-}$ skin only are shown in red, those altered in $Irf6^{S413A,S424A/S413A,S424A}$ only are shown in green and those altered in both genotypes are shown in blue. The Pearson correlation coefficient ($R$ value) is 0.72.

**a** WB: E18.5 skin

**b** K14 K10 Ki67

**c** H&E Stomach

**d**

**e** WB: E18.5 skin

**f**

K14

K10

Ki67

*WT*    *Irf6*^S413E.S424E/S413E.S424E    *Irf6*^S413E.S424E/S413E.S424E *Ripk4*^D161N/D161N    *Irf6*^+/S413E.S424E *Ripk4*^+/D161N    *Irf6*^+/S413E.S424E *Ripk4*^D161N/D161N

**g** H&E

Oral Mucosa

*WT*    *Irf6*^S413E.S424E/S413E.S424E    *Irf6*^S413E.S424E/S413E.S424E *Ripk4*^D161N/D161N    *Irf6*^+/S413E.S424E *Ripk4*^+/D161N    *Irf6*^+/S413E.S424E *Ripk4*^D161N/D161N

**h**

| E18.5 | *Irf6*^+/+ *Ripk4*^1/1 | *Irf6*^+/S413E.S424E *Ripk4*^1/1 | *Irf6*^S413E.S424E/S413E.S424E *Ripk4*^1/1 | *Irf6*^+/+ *Ripk4*^+/D161N | *Irf6*^+/S413E.S424E *Ripk4*^+/D161N | *Irf6*^S413E.S424E/S413E.S424E *Ripk4*^+/D161N | *Irf6*^+/+ *Ripk4*^D161N/D161N | *Irf6*^+/S413E.S424E *Ripk4*^D161N/D161N | *Irf6*^S413E.S424E/S413E.S424E *Ripk4*^D161N/D161N |
|---|---|---|---|---|---|---|---|---|---|
| Observed | 6 | 14 | 7 | 12 | 21 | 12 | 2 | 13 | 4 |
| Expected | 5.7 | 11.4 | 5.7 | 11.4 | 22.8 | 11.4 | 5.7 | 11.4 | 5.7 |

**i**

| E18.5 | *Irf6*^+/+ | *Irf6*^+/S413E.S424E | *Irf6*^S413E.S424E/S413E.S424E |
|---|---|---|---|
| Observed | 13 | 21 | 6 |
| Expected | 10 | 20 | 10 |

**Extended Data Fig. 5** | See next page for caption.

**Extended Data Fig. 5 | The IRF6 phosphomimetic knock-in mouse demonstrates that phosphorylation of IRF6 at Ser413 and Ser424 is essential to prime it for activation. a**, Western blots of wild-type, *Irf6*$^{+/S413A,S424A}$ and *Irf6*$^{S413A,S424A/S413A,S424A}$ E18.5 skin. Representative of three independent experiments. **b**, E18.5 skin sections ($n = 3$ wild type and $n = 3$ *Irf6*$^{S413A,S424A/S413A,S424A}$) stained with H&E or antibodies against K10, K14 and Ki67. Scale bars, 50 μm. **c**, E18.5 sections ($n = 3$ wild type and $n = 3$ *Irf6*$^{S413A,S424A/S413A,S424A}$) stained with H&E, showing fusion of the stratified squamous portion of the stomach in the *Irf6*$^{S413A,S424A/S413A,S424A}$ embryos. Scale bars, 100 μm. **d**, Graph indicates activation of an IRF-responsive luciferase reporter gene at 24 h after transfection of 293T cells with the indicated IRF6 and RIPK4 constructs. IRF6 activity is displayed as fold activity over reporter only. Data are mean ± s.d. ($n = 5$). IRF6(R84C) is a DNA-binding mutant, and represents a negative control in this experiment. Unpaired two-tailed Fisher's exact test with 95%

confidence interval. **e**, Western blots of wild-type, *Irf6*$^{S413E,S424E/S413E,S424E}$ and *Irf6*$^{-/-}$ E18.5 skin. Representative of three independent experiments. **f**, E18.5 skin sections from wild-type, *Irf6*$^{S413E,S424E/S413E,S424E}$, *Irf6*$^{S413E,S424E/S413E,S424E}$*Ripk4*$^{D161N/D161N}$, *Irf6*$^{S413E,S424E/S413E,S424E}$*Ripk4*$^{+/D161N}$ and *Irf6*$^{+/S413E,S424E}$*Ripk4*$^{D161N/D161N}$ embryos ($n = 3$ each) stained with antibodies against K10, K14 and Ki67. Scale bars, 50 μm. **g**, E18.5 sections from wild-type, *Irf6*$^{S413E,S424E/S413E,S424E}$, *Irf6*$^{S413E,S424E/S413E,S424E}$*Ripk4*$^{D161N/D161N}$, *Irf6*$^{S413E,S424E/S413E,S424E}$*Ripk4*$^{+/D161N}$ and *Irf6*$^{+/S413E,S424E}$*Ripk4*$^{D161N/D161N}$ embryos ($n = 3$ each) stained with H&E, showing fusion of the squamous epithelium at the mouth and fusion of the tongue to the palate. Scale bars, 100 μm. **h**, Table showing the observed and expected numbers at E18.5 of offspring that were generated from intercrossing *Ripk4*$^{+/D161N}$*Irf6*$^{+/S413E,S424E}$ mice. **i**, Table showing the observed and expected numbers at E18.5 of offspring that were generated from intercrossing *Irf6*$^{+/S413E.S424E}$ mice.
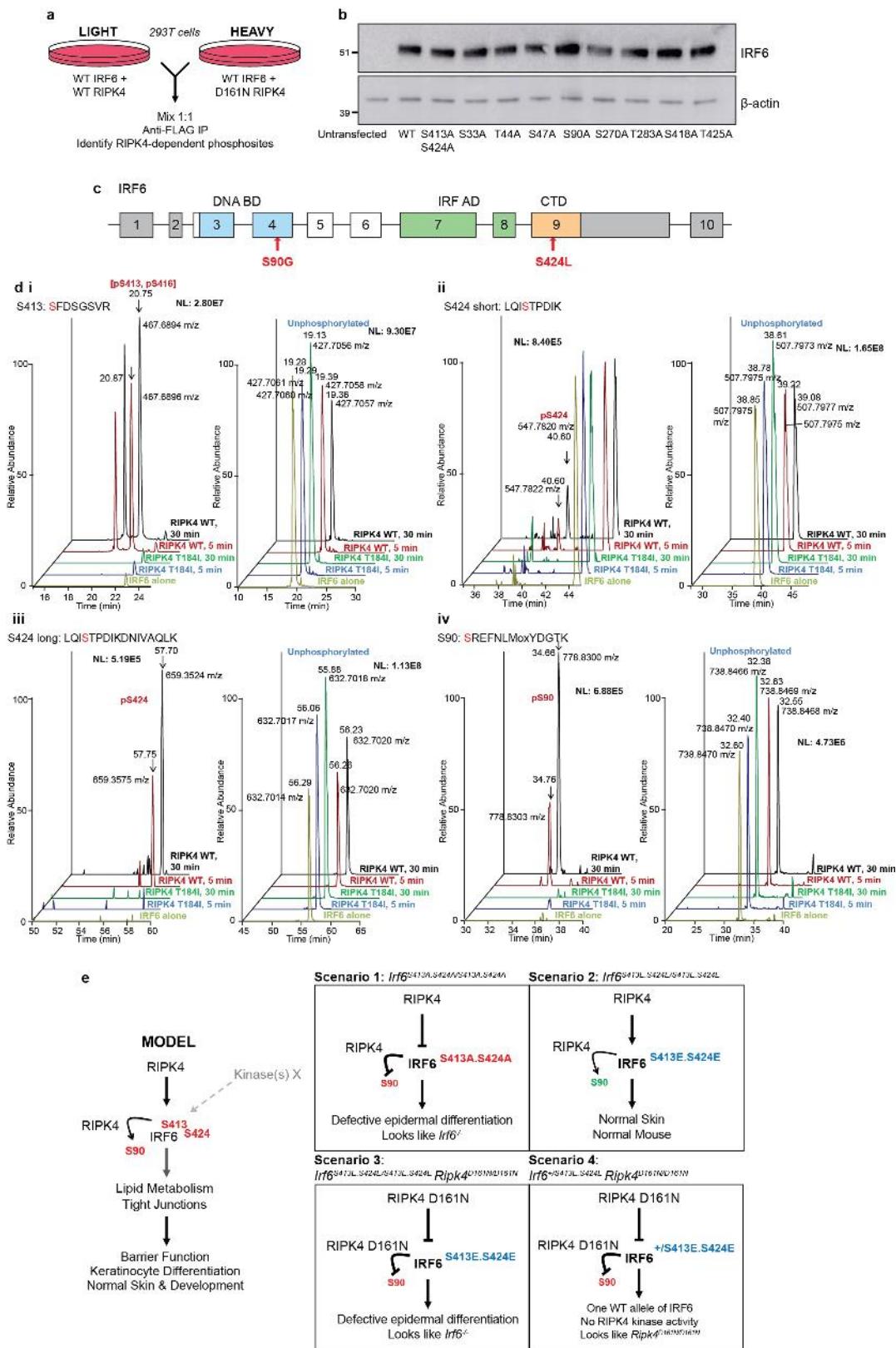
**Extended Data Fig. 6** | See next page for caption.

**Extended Data Fig. 6 | SILAC identifies Ser90 as an additional IRF6 phosphorylation site that is essential for its activation in vitro.**
**a**, Schematic of the SILAC experiment that was performed to identify additional RIPK4-dependent phosphorylation sites on IRF6. **b**, Western blots of 293T cells expressing the IRF6 phosphorylation mutants that were used in luciferase reporter assays. Representative of three independent experiments. **c**, Schematic of the human *IRF6* locus. Exons are shown as rectangles, introns as interconnecting lines and untranslated regions are shaded in grey. The DNA-binding domain (DNA BD), IRF-association domain (IRF AD) and C-terminal domain (CTD) are highlighted in blue, green and orange, respectively. The relative positions of two patient mutations, S90G (which gives rise to VWS) and S424L (which gives rise to PPS) are displayed in red. The locus is drawn approximately to scale. **d**, Extracted ion chromatograms of phosphorylated peptides and their unmodified counterparts at pS(413,416) (SFDSGSVR) (**i**), pS424 (short peptide: LQISTPDIK) (**ii**), pS424 (long peptide: LQISTPDIKDNIVAQLK) (**iii**) and pS90 (SREFNLMoxYDGTK) (**iv**). Recombinant full-length IRF6 was incubated with the RIPK4 kinase domain (either wild type or T184I (a BPS mutation that produces a kinase-dead version of RIPK4) for 5 and 30 min. No phosphorylation was observed when IRF6 was incubated with

RIPK4(T184I). Representative of two experiments. **e**, Proposed model for IRF6 regulation by RIPK4. RIPK4 (or kinase(s) X) phosphorylates IRF6 at Ser413 and Ser424, which act as 'priming' sites. Priming enhances the phosphorylation of IRF6 by RIPK4 at an additional site that is essential for IRF6 activation, Ser90. This allows normal skin differentiation and development. In the first scenario, in which *Irf6* is mutated to $Irf6^{S413,S424A/S413A,S424A}$, RIPK4 (or other kinases) cannot phosphorylate Ser413 and Ser424. IRF6 is thus non-functional, so the $Irf6^{S413A,S424A/S413A,S424A}$ knock-in mouse phenocopies the $Irf6^{-/-}$ mouse. In the second scenario, in which *Irf6* is mutated to $Irf6^{S413E,S424E/S413E,S424E}$, the Glu residues at Ser413 and Ser424 mimic priming and allow RIPK4 to phosphorylate IRF6 at Ser90. Thus, IRF6 is functional, and $Irf6^{S413E,S424E}$ resembles the wild type. In the third, double-mutant scenario ($Irf6^{S413E,S424E/S413E,S424E}Ripk4^{D161N/D161N}$), despite effective IRF6 priming at Ser413 and Ser424 as a result of Glu substitutions, RIPK4 is kinase-dead and therefore cannot phosphorylate IRF6 at Ser90 and activate it. Thus, IRF6 is non-functional and this double mutant phenocopies the $Irf6^{-/-}$ mouse. In the final scenario ($Irf6^{+/S413E,S424E}Ripk4^{D161N/D161N}$), one wild-type allele of *Irf6* is present (sufficient for normal IRF6 function); however, there is no RIPK4 kinase activity. Therefore, these mice phenocopy $Ripk4^{D161N/D161N}$ mice.
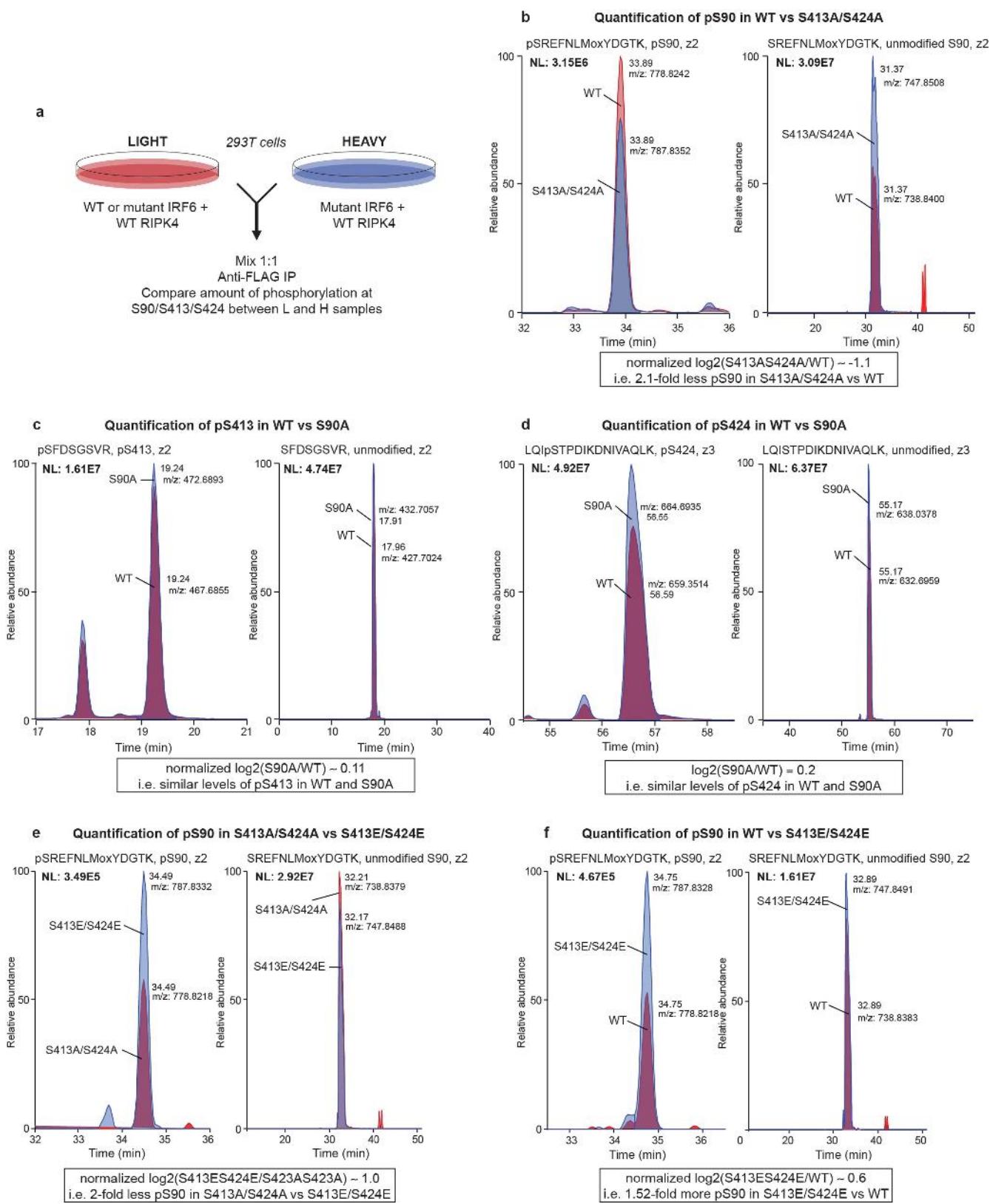
**Extended Data Fig. 7 |** See next page for caption.

**Extended Data Fig. 7 | SILAC experiments demonstrate that mutating IRF6 Ser413 and Ser424 to Ala reduces the amount of phosphorylation at IRF6 Ser90 by half. a**, Schematic of the SILAC experiment that was performed to compare the amount of phosphorylation at S90, S413 and S424 between wild-type and mutant IRF6. **b**, Extracted ion chromatograms of the peptide phosphorylated at Ser90 (pSREFNLMoxYDGTK) and its unmodified counterpart (SREFNLMoxYDGTK), from wild-type IRF6 (light labelled) and IRF6(S413A/S424A) (heavy labelled). More phosphorylation at Ser90 was observed in wild-type IRF6 than in IRF6(S413A/S424A) ($\log_2$(S413A. S424A/WT) = −0.5) even though there was more total IRF6(S413A/ S424A) ($\log_2$(S413A.S424A/WT) = 0.6). When normalized for total IRF6 levels, $\log_2$(S413A.S424A/WT) ≈ −1.1; that is, 2.1-fold less pS90 in IRF6(S413A/S424A) compared to wild-type IRF6. **c**, Extracted ion chromatograms of the peptide phosphorylated at Ser413 or Ser416 (pSFDSGSVR) and its unmodified counterpart (SFDSGSVR), from wild-type IRF6 (light labelled) and IRF6(S90A) (heavy labelled). When normalized for total IRF6 levels, $\log_2$(S90A/WT) = 0.11, so pS413 is at a similar level in wild-type IRF6 and IRF6(S90A). **d**, Extracted ion chromatograms of the peptide phosphorylated at Ser424 (long peptide: LQIpSTPDIKDNIVAQLK) and its unmodified counterpart (long peptide: LQISTPDIKDNIVAQLK), from wild-type IRF6 (light labelled) and IRF6(S90A) (heavy labelled). pS424 is at a similar level in wild-type IRF6

and IRF6(S90A) ($\log_2$(S90A/WT) = 0.2). **e**, Extracted ion chromatograms of the peptide phosphorylated at Ser90 (pSREFNLMoxYDGTK) and its unmodified counterpart (SREFNLMoxYDGTK), from IRF6(S423A/ S424A) (light labelled) and IRF6(S413E/S424E) (heavy labelled). More pS90 was observed in IRF6(S413E/S424E) than in IRF6(S413A/S424A) ($\log_2$(S413E.S424E/S413A.S424A) = 0.6) even though there was slightly less total IRF6(S413E/S424E) ($\log_2$(S413E.S424E/S413A.S424A) = −0.4). When normalized for total IRF6 levels, $\log_2$(S413E.S424E/S413A. S424A) ≈ 1.0; that is, twofold less pS90 in IRF6(S413A/S424A) compared to IRF6(S413E/S424E). **f**, Extracted ion chromatograms of the peptide phosphorylated at Ser90 (pSREFNLMoxYDGTK) and its unmodified counterpart (SREFNLMoxYDGTK), from wild-type IRF6 (light labelled) and IRF6(S413E/S424E) (heavy labelled). More pS90 was observed in IRF6(S413E/S424E) than in wild-type IRF6 ($\log_2$(S413E.S424E/ WT) = 0.7) even though there was also slightly more total IRF6(S413E/ S424E) ($\log_2$(S413E.S424E/WT) = 0.1). When normalized for total IRF6 levels, $\log_2$(S413E.S424E/WT) ≈ 0.6; that is, 1.52-fold more pS90 in IRF6(S413E/S424E) compared to wild-type IRF6. Chromatograms are representative of two experiments. Notably, these are extracted ion chromatograms of representative peptide spectral matches (PSMs). The reported $\log_2$ values were calculated from all the different types of PSMs that cover the phosphorylation sites of interest.
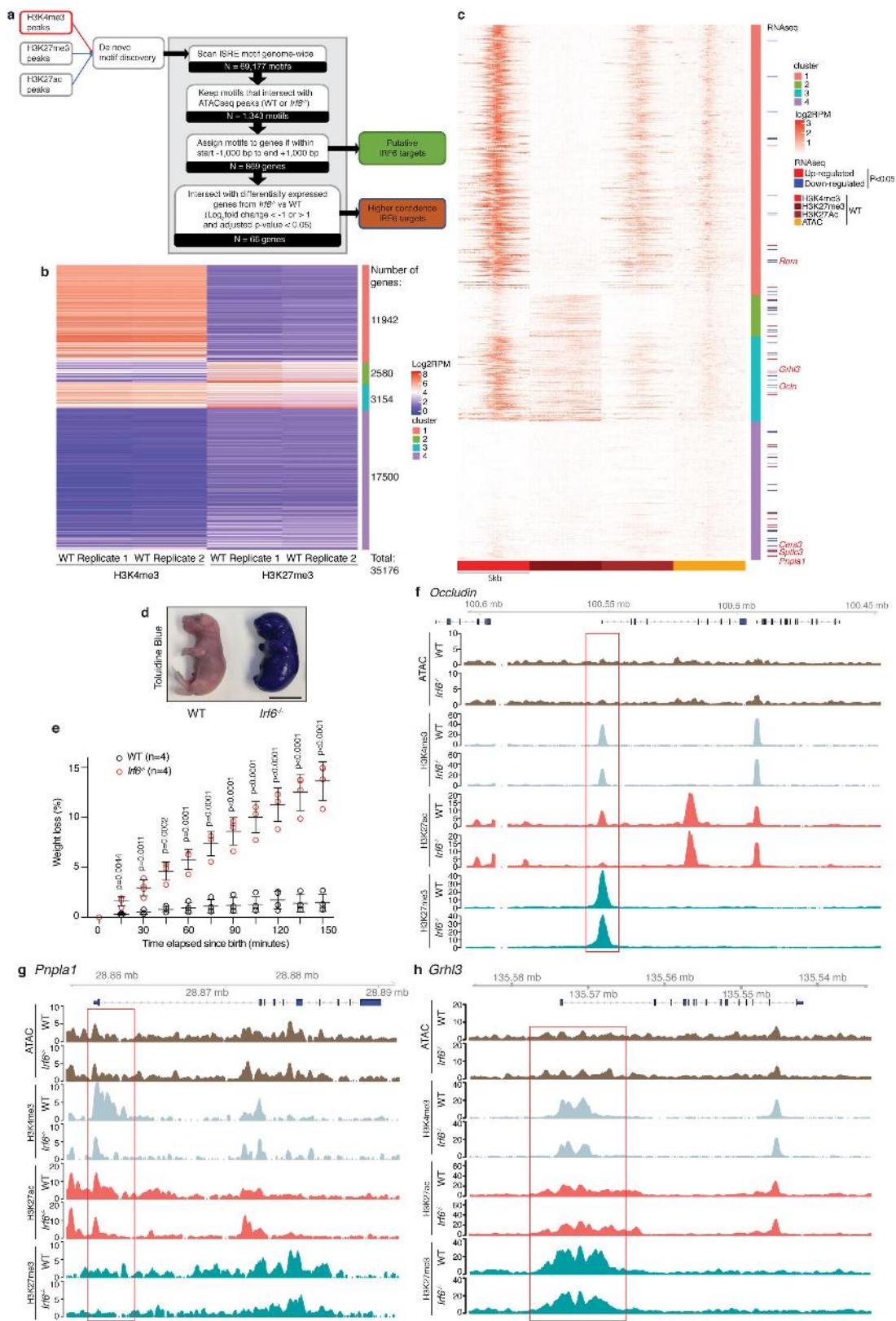
**Extended Data Fig. 8** | See next page for caption.

**Extended Data Fig. 8 | Bioinformatic analysis showing that IRF6 target genes are involved in lipid metabolism and tight-junction formation, which are essential for epidermal barrier function. a**, Schematic of the analysis pipeline that was used to identify putative and high-confidence IRF6 targets. The histone ChIP–seq dataset that yielded an IRF6 motif (the ISRE) is marked in red. **b**, Heat map showing clustering of 35,176 genes into four clusters: active, repressed, bivalent and low-signal promoter groups (top to bottom) ($n = 11,942$, $n = 2,580$, $n = 3,154$ and $n = 17,500$, respectively). Each row reports $\log_2$-transformed reads per million (RPM) in a 5,000-bp window around the start of the gene for H3K4me3 (left two columns) and H3K27me3 (right two columns) in the wild-type background. **c**, Distribution of $\log_2$-transformed read counts in a 5,000-bp window (binned at 50 bp) around 815 putative IRF6 targets (54 out of 869 targets were filtered out owing to low signal) for H3K4me3, H3K27me3, H3K27ac and ATAC-seq signal in the wild-type background, sorted in the

same way as **b**. The number of putative IRF6 targets in active, repressed, bivalent and low-signal groups are 412, 64, 129 and 210, respectively. Genes that are significantly up- or downregulated in $Irf6^{-/-}$ versus wild-type E16.5 skin ($n = 3$ each, from the RNA-seq dataset) are shown in red and blue to the right. $P$ values were obtained using a moderated $t$-test (two-sided) with the Benjamini–Hochberg method for multiple comparisons. **d**, Toluidine blue staining of wild-type and $Irf6^{-/-}$ E18.5 embryos ($n = 3$ each). Scale bar, 1 cm. **e**, Graph showing weight loss of wild-type control ($n = 4$) and $Irf6^{-/-}$ ($n = 4$) newborn mice over time. Data are mean ± s.d. Unpaired, two-tailed Fisher's exact test with 95% confidence interval. **f–h**, Screenshots showing normalized read coverage for ATAC-seq, H3K4me3, H3K27ac and H3K27me3 in wild-type and $Irf6^{-/-}$ skin for occludin ($Ocln$) (**e**), $Pnpla1$ (**f**) and $Grhl3$ (**g**). Genes are indicated in blue in the top panel; red boxes highlight the signal at the start of the gene. Representative of $n = 2$ for wild type and $Irf6^{-/-}$.
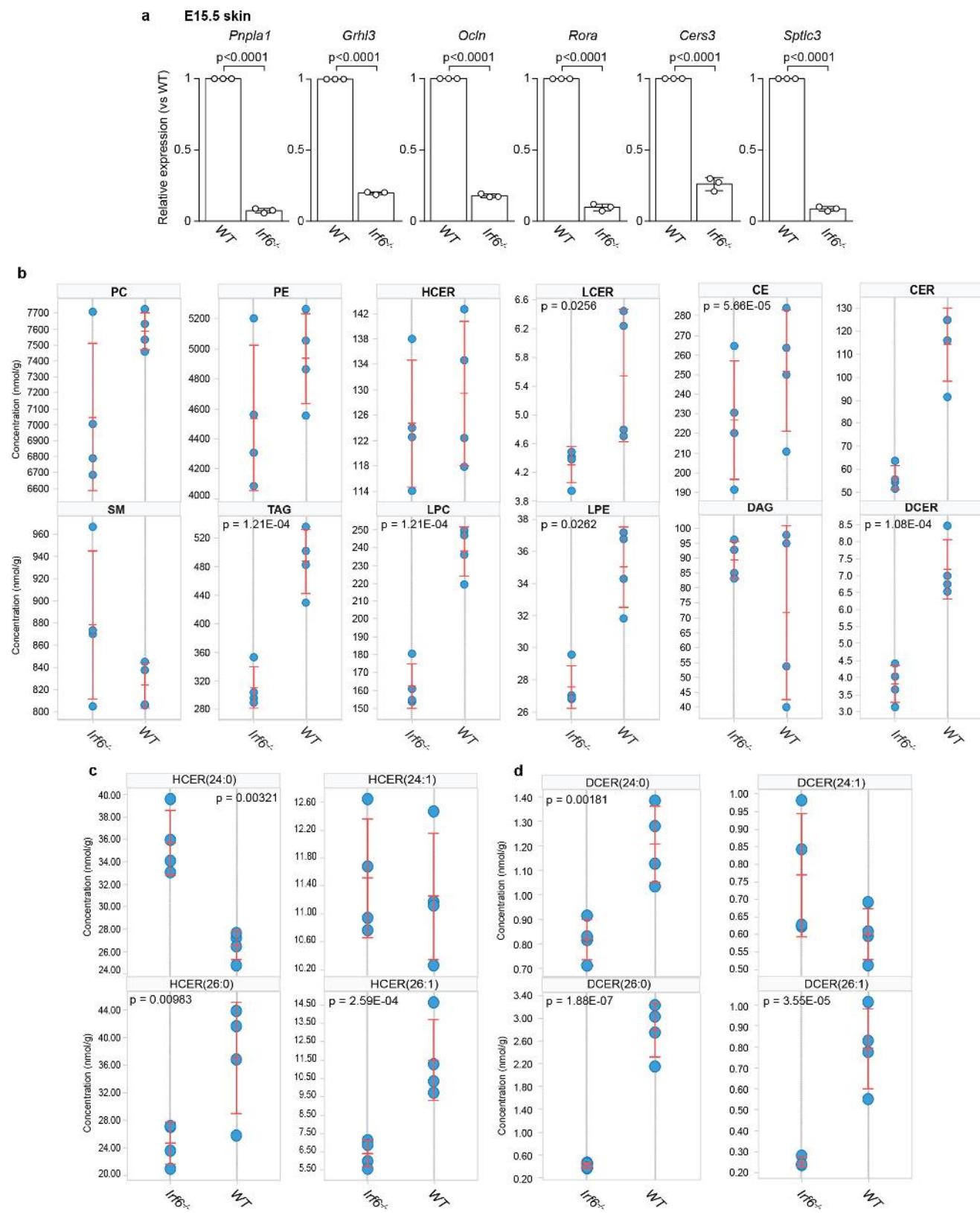
**Extended Data Fig. 9 |** See next page for caption.

**Extended Data Fig. 9 | Lipid profiling showing a decrease in the quantity of certain classes of lipids in $Irf6^{-/-}$ compared to wild-type skin. a**, Graphs of relative mRNA expression of $Grhl3$, $Ocln$, $Pnpla1$, $Sptlc3$, $Cers3$ and $Rora$ in E15.5 skin. Each circle represents a different embryo, $n = 3$. Centre represents the mean; error bars denote s.d. Unpaired, two-tailed Fisher's exact test with 95% confidence interval. **b**, Quantification of 12 classes of lipids in wild-type and $Irf6^{-/-}$ E16.5 skin ($n = 4$ each) (nmol per g of wet skin). Data are mean ± s.d. FDRs (or adjusted $P$ value) of <0.05 are shown on the graph. $P$ values were obtained using a moderated $t$-test (two-sided) with the Benjamini–Hochberg method for multiple comparisons. **c**, Quantification of four different species of HCERs in wild-type and $Irf6^{-/-}$ E16.5 skin ($n = 4$ each) (nmol per g of wet skin). Each dot

represents one skin sample. Centre represents the mean; error bars denote s.d. FDRs of <0.05 are labelled on the graph. $P$ values were obtained using a moderated $t$-test (two-sided) with the Benjamini–Hochberg method for multiple comparisons. There is a significant decrease in species with ultra-long-chain acyl moieties ($\geq$C26) in $Irf6^{-/-}$ compared to wild-type skin. **d**, Quantification of four different species of DCERs in wild-type and $Irf6^{-/-}$ E16.5 skin ($n = 4$ each) (nmol per g of wet skin). Each dot represents one skin sample. Centre represents the mean; error bars denote s.d. FDRs of <0.05 are labelled on the graph. $P$ values were obtained using a moderated $t$-test (two-sided) with the Benjamini–Hochberg method for multiple comparisons. There is a significant decrease in species with ultra-long-chain acyl moieties ($\geq$C26) in $Irf6^{-/-}$ compared to wild-type skin.
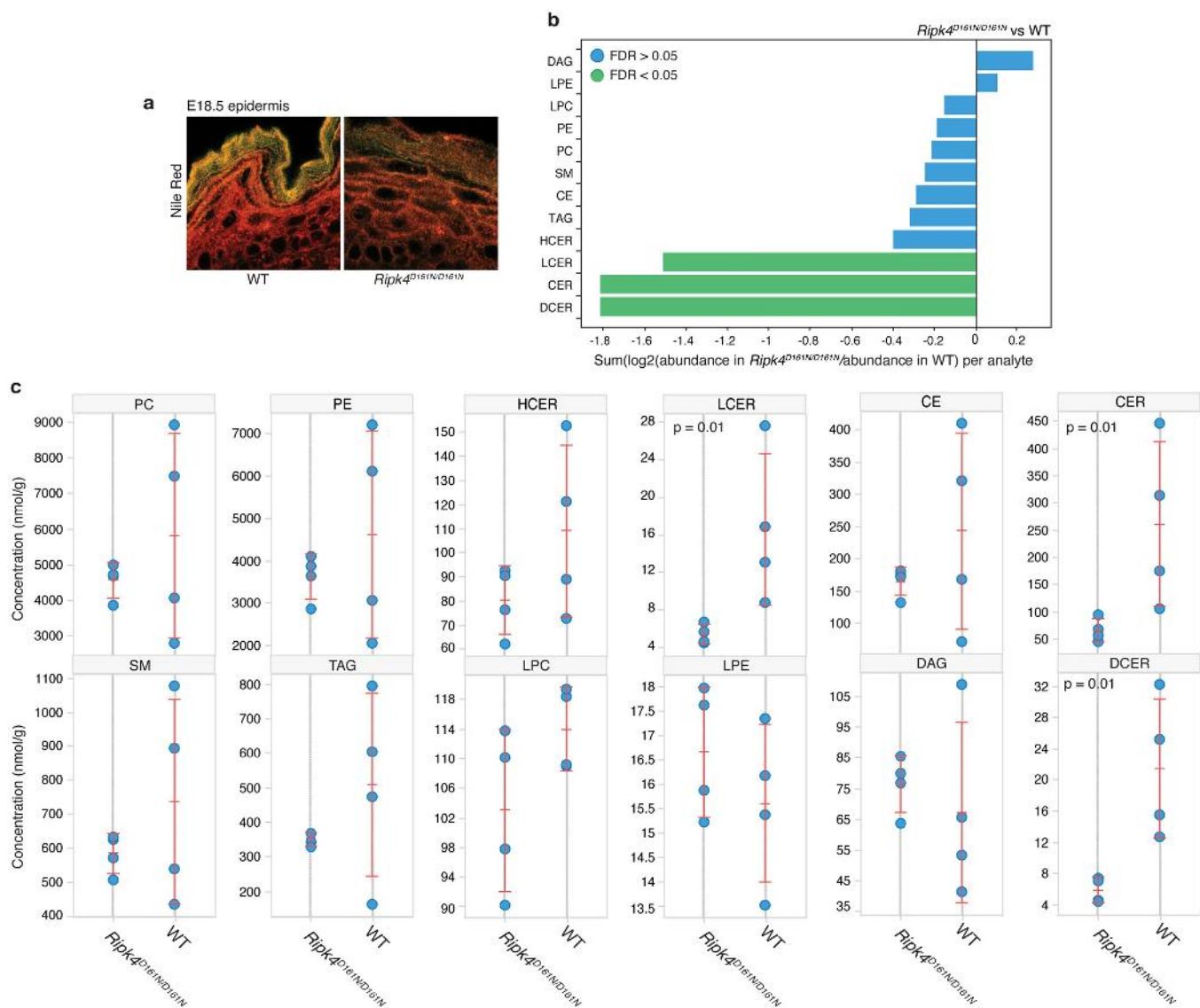
**Extended Data Fig. 10 | Lipid profiling showing a decrease in the quantity of certain classes of lipids in $Ripk4^{D161N/D161N}$ compared to wild-type skin.** **a**, E18.5 skin sections from wild-type and $Ripk4^{D161N/D161N}$ embryos ($n = 3$) were stained with Nile red fluorescent dye, which indicates polar lipids in red and non-polar lipids in green. $72.66 \times 72.66$ microns. **b**, Profiling of 12 classes of lipids in wild-type and $Ripk4^{D161N/D161N}$ E16.5 skin. The $x$ axis denotes the sum(log$_2$(abundance in $Ripk4^{D161N/D161N}$/abundance in wild type)) per analyte; values less than 0 indicate a decrease, and values greater than 0 an increase, in $Ripk4^{D161N/D161N}$ versus wild-type skin. Green bars denote a significant FDR of <0.05.

Moderated $t$-test (two-sided). $P$ values were obtained using a moderated $t$-test (two-sided) with the Benjamini–Hochberg method for multiple comparisons. There are significantly lower levels of CERs, DCERs and LCERs in $Ripk4^{D161N/D161N}$ compared to wild-type skin. Levels of the other classes of lipids are unchanged. **c**, Quantification of 12 classes of lipids in wild-type and $Ripk4^{D161N/D161N}$ E16.5 skin ($n = 4$) (nmol per g of wet skin). Each dot represents one skin sample. Centre represents the mean; error bars denote s.d. FDRs of <0.05 are shown on the graph. $P$ values were obtained using a moderated $t$-test (two-sided) with the Benjamini–Hochberg method for multiple comparisons.

# nature research

| | |
|---|---|
| Corresponding author(s): | Vishva M Dixit |
| Last updated by author(s): | Aug 19, 2019 |

# Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist .

## Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

| n/a | Confirmed | |
|---|---|---|
| ☐ | ☒ | The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement |
| ☐ | ☒ | A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly |
| ☐ | ☒ | The statistical test(s) used AND whether they are one- or two-sided<br>*Only common tests should be described solely by name; describe more complex techniques in the Methods section.* |
| ☒ | ☐ | A description of all covariates tested |
| ☐ | ☒ | A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons |
| ☐ | ☒ | A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals) |
| ☐ | ☒ | For null hypothesis testing, the test statistic (e.g. $F$, $t$, $r$) with confidence intervals, effect sizes, degrees of freedom and $P$ value noted<br>*Give P values as exact values whenever suitable.* |
| ☒ | ☐ | For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings |
| ☒ | ☐ | For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes |
| ☐ | ☒ | Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated |

*Our web collection on statistics for biologists contains articles on many of the points above.*

## Software and code

Policy information about availability of computer code

| | |
|---|---|
| Data collection | Embryo immunofluorescence skin images were acquired with Leica Application Suite v4.11.0.  Immunohistochemistry and histology images were acquired with Leica Application Suite v4.6.0. RNA sequencing was performed on an Illumina HiSeq4000 (Illumina). ChIP-seq and ATAC-seq were performed on a NextSeq 500 sequencer (Illumina). Peptides for Mass Spectrometry were analyzed using an LTQ Orbitrap Elite mass spectrometer (ThermoFisher, San Jose, CA). Lipids were analyzed using a SelexION enabled 6500 QTRAP (Sciex, Redwood City, CA). |
| Data analysis | RNA-seq, ATAC-seq and ChIP-seq data were analyzed with the help of the following tools:<br>1) GSNAP (version 2013-11-01; research-pub.gene.com/gmap/)<br>2) R version 3.4.3<br>3) Limma R package version 3.34.9<br>4) DESeq2 R package version 1.18.1<br>5) Gviz R package version 1.22.2<br>6) Macs2 version 2.1.0<br>7) HOMER version 4.7<br>8) FIMO version 5.0.4<br>9) BEDTools version 2.22.1<br>10) EGSEA R package version 1.6.1<br>11) ENCODE pipeline with minor modifications (https://www.encodeproject.org/atac-seq/#standards)<br><br>Graphs were generated with Prism 6.<br>Lipidomics data were uploaded to a Spotfire (TIBCO Software, Sommerville, MA) dashboard for interrogation and plot generation.<br>LC-MS/MS statistical analysis and data visualization was performed using R (version 3.5.0). |

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

## Data

Policy information about <u>availability of data</u>

All manuscripts must include a <u>data availability statement</u>. This statement should provide the following information, where applicable:
- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

The datasets generated during and/or analysed during the current study are available from the corresponding authors on reasonable request.
To access RNA-seq, ATAC-seq and ChIP-seq data in GEO go to: https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE124067
password for access is: gnynyeienrknpol

# Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

[✖] Life sciences  [ ] Behavioural & social sciences  [ ] Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

# Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

| | |
|---|---|
| Sample size | Sample sizes were determined empirically, no sample size calculations were performed. Whenever possible, at least 3 animals per genotype were analyzed to be sure differences were reproducible, while minimizing the number of animals used per experiment per the 3Rs. Variability in the ex vivo assays used in this study tends to be very low because the phenotypes are so strong, so n=3 is the accepted norm in the field. |
| Data exclusions | No data were excluded. |
| Replication | Whenever possible, readouts were performed with at least 3 animals of a given genotype and all attempts at replication were successful. |
| Randomization | Mice were grouped according to genotype, not randomized. Randomization was not necessary because there were no experiments involving treatment groups. |
| Blinding | Embryo genotypes were unknown when animals were harvested but investigators were not blinded to outcome assessment. Blinding was not necessary as there was no quantification or analysis of subtle phenotypes. |

# Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

### Materials & experimental systems

| n/a | Involved in the study |
|---|---|
| [ ] | [✖] Antibodies |
| [ ] | [✖] Eukaryotic cell lines |
| [✖] | [ ] Palaeontology |
| [ ] | [✖] Animals and other organisms |
| [✖] | [ ] Human research participants |
| [✖] | [ ] Clinical data |

### Methods

| n/a | Involved in the study |
|---|---|
| [ ] | [✖] ChIP-seq |
| [✖] | [ ] Flow cytometry |
| [✖] | [ ] MRI-based neuroimaging |

## Antibodies

| | |
|---|---|
| Antibodies used | For WB the following antibodies were used: anti-FLAG (Sigma; used at 1:5000), anti-MYC (GeneTex; used at 1:2000), anti-IRF6 (rabbit mAb, GEN168NP-F1, Genentech; used at 1 μg/ml), anti-RIPK4 (rat mAb, 3E9.3.1; Genentech, used at 1 μg/ml) and anti-β-Actin (CST, 13E5, used at 1:2000). |
| | For IHC the following antibodies were used: anti-Cytokeratin-14 (PRB-155P, Biolegend, 0.1 μg/ml), rabbit anti-Cytokeratin-10 (PRB-159P, Biolegend, 1 μg/ml), mouse anti-Ki67 (clone SP6, RM-9106-S, ThermoFisher, 1:200), or rabbit anti-IRF6 (GEN168NPD5, Genentech, 0.5 μg/ml). |
| | ChIP was performed with validated antibodies against H3K27me3 (Active Motif 39155), H3K27Ac (Active Motif 39133) or H3K4me3 (Active Motif 39159). |

| Validation | anti-IRF6 (rabbit mAb, GEN168NP-F1, Genentech; used at 1 µg/ml) was validated for WB using over-expression of mouse proteins in 293T cells in Huang, C. S. et al. Crystal Structure of Ripk4 Reveals Dimerization-Dependent Kinase Activity. Structure (2018). doi:10.1016/j.str.2018.04.002.<br>anti-IRF6 (GEN168NPD5, Genentech, 0.5 µg/ml) was validated for IHC using WT and Irf6-/- mouse tissue.<br>anti-RIPK4 (rat mAb, 3E9.3.1; Genentech, used at 1 µg/ml) was validated for WB using WT and Ripk4-/- mouse tissue.<br>Validation data for all commercial antibodies are available on vendor websites. |
|---|---|

## Eukaryotic cell lines

Policy information about cell lines

| Cell line source(s) | 293T GNE from Genentech |
|---|---|
| Authentication | Short Tandem Repeat (STR) profiles were determined using the Promega PowerPlex 16 System. This was performed once and compared to external STR profiles of cell lines to determine cell line ancestry.<br>Source: GNE, Tracking ID: 129641, Cell Line Name: 293T GNE, Number of Markers: 16, D16S539: 9, 13, D18S51: 17, 19, D2S1338: n/a, D5S818: 8, 9, D7S820: 11, 12, vWA: 16, 19, TPOX: 11, TH01: 7, 9.3, AMEL: X, FGA: 20, 23, D3S1358: 15, 17, CSF1PO: 11, 12, D13S317: 12, D8S1179: 11, 12, 14, 15, Penta D: 9, 10, Penta E: 7, 14, 15, D19S433: n/a, D21S11: 28, 30.2<br>SNP profiles were performed each time new stocks were expanded for cryopreservation. Cell line identity was verified by high-throughput SNP profiling using Fluidigm multiplexed assays. SNPs were selected based on minor allele frequency and presence on commercial genotyping platforms. SNP profiles were compared to SNP calls from available internal and external data to determine or confirm ancestry. SNPs analyzed: rs11746396, rs16928965, rs2172614, rs10050093, rs10828176, rs16888998, rs16999576, rs1912640, rs2355988, rs3125842, rs10018359, rs10410468, rs10834627, rs11083145, rs11100847, rs11638893, rs12537, rs1956898, rs2069492, rs10740186, rs12486048, rs13032222, rs1635191, rs17174920, rs2590442, rs2714679, rs2928432, rs2999156, rs10461909, rs11180435, rs1784232, rs3783412, rs10885378, rs1726254, rs2391691, rs3739422, rs10108245, rs1425916, rs1325922, rs1709795, rs1934395, rs2280916, rs2563263, rs10755578, rs1529192, rs2927899, rs2848745, rs10977980 |
| Mycoplasma contamination | All stocks are tested for mycoplasma prior to and after cells are cryopreserved.<br>Two methods are used to avoid false positive/negative results: Lonza Mycoalert and Stratagene Mycosensor. |
| Commonly misidentified lines<br>(See ICLAC register) | No commonly misitentified cell lines were used. |

## Animals and other organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research

| Laboratory animals | All mice (Mus musculus) were on a C57BL/6N genetic background. Timed matings were set up with males and females from 6 to 26 weeks of age and E15.5-E18.5 embryos and P0 pups were analysed (both males and females). Mutant strains included Ripk4 +/D161N, Ripk4 +/flox, Irf6 +/flox, Irf6 +/S413A.S424A, Irf6 +/S413E.S424E (all generated in this study) and K14-Cre transgene (Dassule et al., 2000) |
|---|---|
| Wild animals | This study did not involve wild animals. |
| Field-collected samples | This study did not involve samples collected from the field. |
| Ethics oversight | All mouse studies were approved by the Genentech institutional animal care and use committee (IACUC). |

Note that full information on the approval of the study protocol must also be provided in the manuscript.

## ChIP-seq

### Data deposition

☒ Confirm that both raw and final processed data have been deposited in a public database such as GEO.

☒ Confirm that you have deposited or provided access to graph files (e.g. BED files) for the called peaks.

| Data access links<br>*May remain private before publication.* | To access data in GEO go to: https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE124067<br>password for access is: gnynyeienrknpol |
|---|---|
| Files in database submission | fastq and peak BED files |
| Genome browser session<br>(e.g. UCSC) | N/A |

### Methodology

| Replicates | 2 replicates per genotype, WT and Irf6-/- |
|---|---|

| | |
|---|---|
| Sequencing depth | Sample name Total reads Total mapped reads Unique reads<br>WT1_H3K27Ac 42298785 38250648 36507704<br>WT1_H3K27me3 40554967 32745471 30618687<br>WT1_H3K4me3 33811247 30322682 29081425<br>WT2_H3K27Ac 39604330 34131965 32357247<br>WT2_H3K27me3 44051505 35671630 32850109<br>WT2_H3K4me3 32546866 29884441 28515422<br>Irf6-KO1_H3K27Ac 52643008 40551078 38434385<br>Irf6-KO1_H3K27me3 44209789 35320541 32759367<br>Irf6-KO1_H3K4me3 36892069 23316940 21988746<br>Irf6-KO2_H3K27Ac 43889226 35147138 33433279<br>Irf6-KO2_H3K27me3 43928544 36829044 34687925<br>Irf6-KO2_H3K4me3 40490371 36859831 35198045<br>Pooled_Input 35956589 34334457 30100724<br>All reads are single-end reads with read length of 75bp. |
| Antibodies | ChIP was performed with validated antibodies against H3K27me3 (Active Motif 39155), H3K27Ac (Active Motif 39133) or H3K4me3 (Active Motif 39159). |
| Peak calling parameters | GSNAP version 2013-11-01 was used to map raw FASTQ reads to mouse genome (GRCm38/mm10) using the following parameters: -M 2 -n 10 -B 2 -i 1 --pairmax-dna=1000 --terminalthreshold=1000 --gmap-mode=none --clip-overlap<br>Macs2 commands are listed below:<br>macs2 callpeak -B --SPMR -p 1e-7 -t WT1_H3K27Ac.analyzed.bam --c Pooled_Input.analyzed.bam -f BAM -g mm -n WT1_H3K27Ac --outdir macs_out<br>macs2 callpeak -B --SPMR -p 1e-7 -t WT1_H3K4me3.analyzed.bam --c Pooled_Input.analyzed.bam -f BAM -g mm -n WT1_H3K4me3 --outdir macs_out<br>macs2 callpeak -B --SPMR -p 1e-7 -t WT1_H3K4me3.analyzed.bam --c Pooled_Input.analyzed.bam -f BAM -g mm -n WT1_H3K4me3 --outdir macs_out<br>macs2 callpeak -B --SPMR -p 1e-7 -t WT2_H3K27Ac.analyzed.bam --c Pooled_Input.analyzed.bam -f BAM -g mm -n WT2_H3K27Ac --outdir macs_out<br>macs2 callpeak -B --SPMR -p 1e-7 -t WT2_H3K27me3.analyzed.bam --c Pooled_Input.analyzed.bam -f BAM -g mm -n WT2_H3K27me3 --outdir macs_out<br>macs2 callpeak -B --SPMR -p 1e-7 -t WT2_H3K4me3.analyzed.bam --c Pooled_Input.analyzed.bam -f BAM -g mm -n WT2_H3K4me3 --outdir macs_out<br>macs2 callpeak -B --SPMR -p 1e-7 -t Irf6-KO1_H3K27Ac.analyzed.bam --c Pooled_Input.analyzed.bam -f BAM -g mm -n Irf6-KO1_H3K27Ac --outdir macs_out<br>macs2 callpeak -B --SPMR -p 1e-7 -t Irf6-KO1_H3K27me3.analyzed.bam --c Pooled_Input.analyzed.bam -f BAM -g mm -n Irf6-KO1_H3K27me3 --outdir macs_out<br>macs2 callpeak -B --SPMR -p 1e-7 -t Irf6-KO1_H3K4me3.analyzed.bam --c Pooled_Input.analyzed.bam -f BAM -g mm -n Irf6-KO1_H3K4me3 --outdir macs_out<br>macs2 callpeak -B --SPMR -p 1e-7 -t Irf6-KO2_H3K27Ac.analyzed.bam --c Pooled_Input.analyzed.bam -f BAM -g mm -n Irf6-KO2_H3K27Ac --outdir macs_out<br>macs2 callpeak -B --SPMR -p 1e-7 -t Irf6-KO2_H3K27me3.analyzed.bam --c Pooled_Input.analyzed.bam -f BAM -g mm -n Irf6-KO2_H3K27me3 --outdir macs_out<br>macs2 callpeak -B --SPMR -p 1e-7 -t Irf6-KO2_H3K4me3.analyzed.bam --c Pooled_Input.analyzed.bam -f BAM -g mm -n Irf6-KO2_H3K4me3 --outdir macs_out |
| Data quality | We followed ENCODE guidelines for data quality assessment. Per ENCODE guidelines, the target NRF (nonredundancy fraction) should be >= 0.8 for 10 million reads, all samples in our study passed this criteria. All samples also had a FRiP (Fraction of reads in peaks) enrichment of greater than 1%.<br>Sample name peaks (>=5% FDR, >= 5 fold change)<br>WT1_H3K27Ac 35820<br>WT1_H3K27me3 16290<br>WT1_H3K4me3 21474<br>WT2_H3K27Ac 46508<br>WT2_H3K27me3 12300<br>WT2_H3K4me3 21510<br>Irf6-KO1_H3K27Ac 40138<br>Irf6-KO1_H3K27me3 18169<br>Irf6-KO1_H3K4me3 20923<br>Irf6-KO2_H3K27Ac 39539<br>Irf6-KO2_H3K27me3 23461<br>Irf6-KO2_H3K4me3 22282 |
| Software | Raw FASTQ reads were mapped using GSNAP version, 2013-11-01, Macs2 version 2.1.0 was used to call peaks. Homer version 4.7 was used to look for enriched motifs in the dataset. Genome-wide scan of ISRE motif was performed using FIMO (part of MEME-Suite version 5.0.4). |

# LETTER

# Light-entrained and brain-tuned circadian circuits regulate ILC3s and gut homeostasis

Cristina Godinho-Silva[1,5], Rita G. Domingues[1,5], Miguel Rendas[1], Bruno Raposo[1], Hélder Ribeiro[1], Joaquim Alves da Silva[1,2], Ana Vieira[1], Rui M. Costa[1,3], Nuno L. Barbosa-Morais[4], Tânia Carvalho[4] & Henrique Veiga-Fernandes[1]*

**Group 3 innate lymphoid cells (ILC3s) are major regulators of inflammation, infection, microbiota composition and metabolism[1]. ILC3s and neuronal cells have been shown to interact at discrete mucosal locations to steer mucosal defence[2,3]. Nevertheless, it is unclear whether neuroimmune circuits operate at an organismal level, integrating extrinsic environmental signals to orchestrate ILC3 responses. Here we show that light-entrained and brain-tuned circadian circuits regulate enteric ILC3s, intestinal homeostasis, gut defence and host lipid metabolism in mice. We found that enteric ILC3s display circadian expression of clock genes and ILC3-related transcription factors. ILC3-autonomous ablation of the circadian regulator *Arntl* led to disrupted gut ILC3 homeostasis, impaired epithelial reactivity, a deregulated microbiome, increased susceptibility to bowel infection and disrupted lipid metabolism. Loss of ILC3-intrinsic *Arntl* shaped the gut 'postcode receptors' of ILC3s. Strikingly, light–dark cycles, feeding rhythms and microbial cues differentially regulated ILC3 clocks, with light signals being the major entraining cues of ILC3s. Accordingly, surgically or genetically induced deregulation of brain rhythmicity led to disrupted circadian ILC3 oscillations, a deregulated microbiome and altered lipid metabolism. Our work reveals a circadian circuitry that translates environmental light cues into enteric ILC3s, shaping intestinal health, metabolism and organismal homeostasis.**

ILC3s have been shown to be part of discrete mucosal neuroimmune cell units[2–5], raising the hypothesis that ILC3s may also integrate systemic neuroimmune circuits to regulate tissue integrity and organismic homeostasis. Circadian rhythms rely on local and systemic cues to coordinate mammalian physiology and are genetically encoded by molecular clocks that allow organisms to anticipate and adapt to extrinsic environmental changes[6,7]. The circadian clock machinery consists of an autoregulatory network of feedback loops primarily driven by the activators ARNTL and CLOCK and the repressors PER1–PER3, CRY1 and CRY2, amongst others[6,7].

Analysis of subsets of intestinal ILCs and their bone marrow progenitors revealed that mature ILC3s express high levels of circadian clock genes (Fig. 1a–c, Extended Data Fig. 1a–d). Notably, ILC3s displayed a circadian pattern of *Per1^Venus* expression (Fig. 1b) and transcriptional analysis of ILC3 revealed circadian expression of master clock regulators and ILC3-related transcription factors (Fig. 1c). To test whether ILC3s are regulated in a circadian manner, we investigated whether intestinal ILC3s require intrinsic clock signals. Thus, we interfered with the expression of the master circadian activator *Arntl*. *Arntl^fl* mice were bred to *Vav1^Cre* mice, allowing conditional deletion of *Arntl* in all haematopoietic cells (*Arntl^ΔVav1* mice). Although *Arntl^ΔVav1* mice displayed normal numbers of intestinal natural killer (NK) cells and enteric group 1 and 2 ILCs, gut ILC3s were severely and selectively reduced in these mice when compared to their wild-type littermate controls (Fig. 1d, e, Extended Data Fig. 2a, b). To more precisely define ILC3-intrinsic effects, we generated mixed bone marrow chimaeras by transferring *Arntl*-competent (*Arntl^fl*) or *Arntl*-deficient (*Arntl^ΔVav1*) bone

marrow against a third-party wild-type competitor into alymphoid hosts (Fig. 1f). Analysis of such chimaeras confirmed cell-autonomous circadian regulation of ILC3s, while their innate and adaptive counterparts were unperturbed (Fig. 1g, Extended Data Fig. 2c).
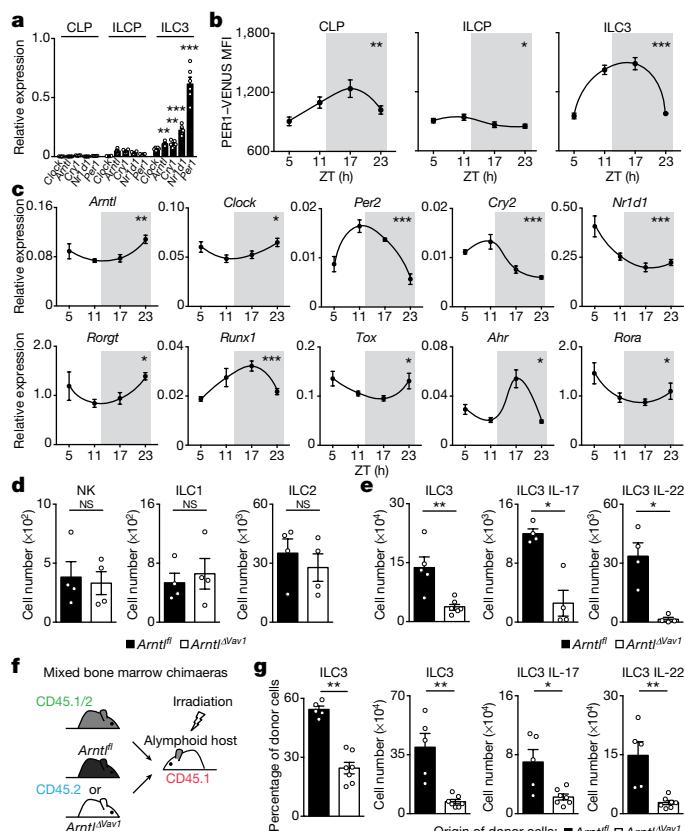


**Fig. 1 | Intestinal ILC3s are controlled in a circadian manner. a**, Gene expression in CLPs, ILCPs and intestinal ILC3s. CLP and ILCP $n = 4$; ILC3 $n = 6$. **b**, PER1–VENUS mean fluorescence intensity (MFI). CLP and ILCP $n = 6$; ILC3 $n = 4$. **c**, Circadian gene expression in enteric ILC3s; $n = 5$. **d**, Intestinal ILC subsets in *Arntl^fl* and *Arntl^ΔVav1* mice; $n = 4$. **e**, Cell numbers of intestinal ILC3s and IL-17- and IL-22-producing ILC3 subsets in *Arntl^fl* and *Arntl^ΔVav1* mice; $n = 4$. **f**, Generation of mixed bone marrow chimaeras. **g**, Percentage of donor cells and cell numbers of ILC3s, IL-17 and IL-22-producing ILC3 subsets in the gut from mixed bone marrow chimaeras. *Arntl^fl* $n = 5$, *Arntl^ΔVav1* $n = 7$. **b, c**, White and grey represent light and dark periods, respectively. Data are representative of three independent experiments. $n$ represents biologically independent samples (**a, c**) or animals (**b, d–g**). Data shown as mean ± s.e.m. **a**, Two-way ANOVA and Tukey's test; **b, c**, cosinor analysis; **d, e, g**, Two-tailed Mann–Whitney $U$ test. *$P < 0.05$; **$P < 0.01$; ***$P < 0.001$; NS, not significant.

[1]Champalimaud Research, Champalimaud Centre for the Unknown, Lisbon, Portugal. [2]Champalimaud Clinical Centre, Champalimaud Centre for the Unknown, Lisbon, Portugal. [3]Zuckerman Mind Brain Behavior Institute, Columbia University, New York, NY, USA. [4]Instituto de Medicina Molecular João Lobo Antunes, Faculdade de Medicina, Universidade de Lisboa, Lisbon, Portugal. [5]These authors contributed equally: Cristina Godinho-Silva, Rita G. Domingues. *e-mail: henrique.veigafernandes@research.fchampalimaud.org
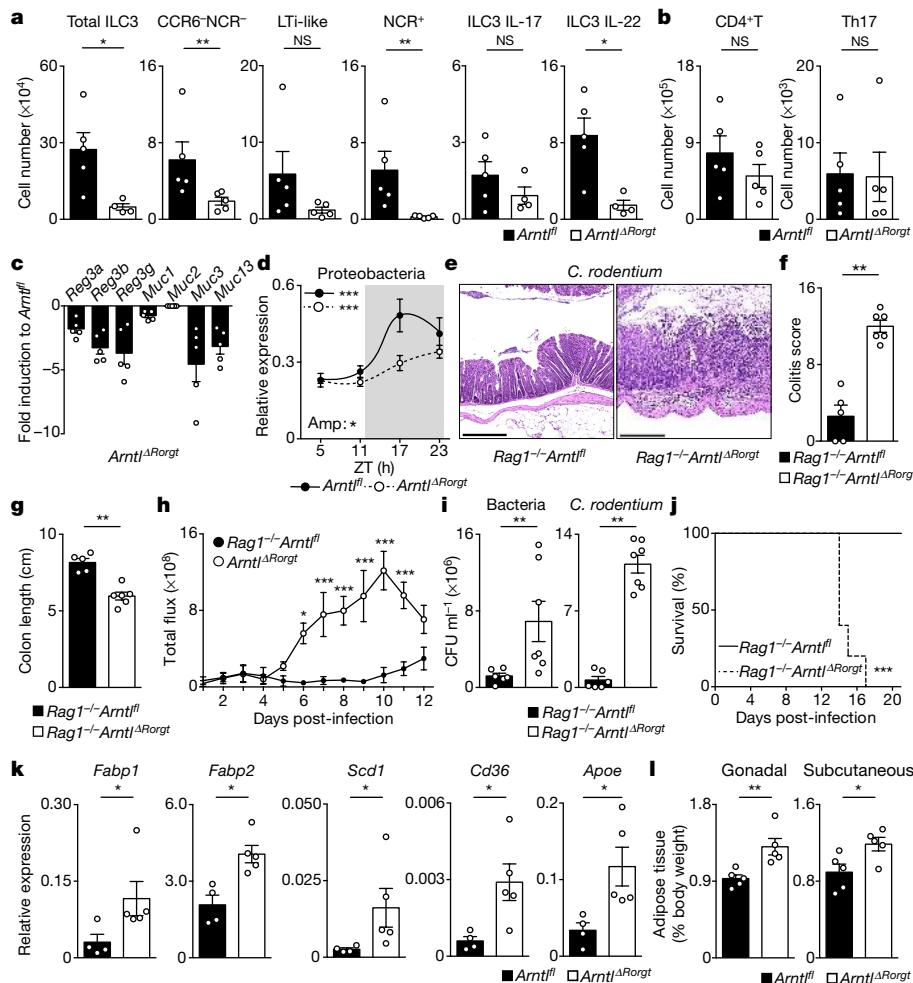
**Fig. 2 | ILC3-intrinsic *Arntl* regulates gut homeostasis and defence.**
**a**, Enteric ILC3s and subtypes in $Arntl^{fl}$ and $Arntl^{\Delta Rorgt}$ mice; $n = 4$.
**b**, Gut T helper cells in $Arntl^{fl}$ and $Arntl^{\Delta Rorgt}$ mice; $n = 5$. **c**, Expression
of epithelial reactivity genes in $Arntl^{\Delta Rorgt}$ mice compared with $Arntl^{fl}$
mice; $n = 5$. **d**, qPCR analysis of Proteobacteria in stools from $Arntl^{fl}$ and
$Arntl^{\Delta Rorgt}$ mice (see Methods). $Arntl^{fl}$ $n = 5$; $Arntl^{\Delta Rorgt}$ $n = 6$. **e–j**, Data
from *C. rodentium*-infected $Rag1^{-/-}Arntl^{fl}$ and $Rag1^{-/-}Arntl^{\Delta Rorgt}$
mice. **e**, Histopathology of colon sections; $n = 5$. **f**, Colitis score; $n = 5$.
**g**, Colon length; $n = 5$. **h**, Infection burden; $Rag1^{-/-}Arntl^{fl}$ $n = 6$,
$Rag1^{-/-}Arntl^{\Delta Rorgt}$ $n = 7$. **i**, Bacterial translocation to the spleen;

$Rag1^{-/-}Arntl^{fl}$ $n = 6$, $Rag1^{-/-}Arntl^{\Delta Rorgt}$ $n = 7$. **j**, Survival; $n = 5$.
**k**, Expression of epithelial lipid transporter genes in $Arntl^{fl}$ ($n = 4$) and
$Arntl^{\Delta Rorgt}$ ($n = 5$) mice. **l**, Gonadal and subcutaneous adipose tissue in
$Arntl^{fl}$ and $Arntl^{\Delta Rorgt}$ mice; $n = 5$. **d**, White and grey represent light and
dark periods, respectively. Scale bars, 250 μm. Data are representative
of at least three independent experiments; $n$ represents biologically
independent animals. Data shown as mean ± s.e.m. **a**, **b**, **f**, **g**, **i**, **k**, Two-
tailed Mann–Whitney $U$ test; **d**, cosinor analysis; **h**, two-way ANOVA
and Sidak's test; **j**, log-rank test; **l**, two-tailed unpaired Student's $t$-test.
*$P < 0.05$; **$P < 0.01$; ***$P < 0.001$; NS, not significant.

To investigate the functional effect of ILC3-intrinsic circadian signals,
we deleted *Arntl* in RORγt-expressing cells by breeding $Rorgt^{Cre}$ mice
(also known as $Rorc^{Cre}$) to $Arntl^{fl}$ mice ($Arntl^{\Delta Rorgt}$ mice). When com-
pared to their wild-type littermate controls, $Arntl^{\Delta Rorgt}$ mice showed
a selective reduction of ILC3 subsets and IL-17- and IL-22-producing
ILC3s (Fig. 2a, b, Extended Data Fig. 3a–j). Notably, independent
deletion of *Nr1d1* also perturbed subsets of enteric ILC3s, further
supporting a role of the clock machinery in ILC3s (Extended Data
Fig. 4a–e). ILC3s have been shown to regulate the expression of genes
related to epithelial reactivity and microbial composition[1]. Analysis
of $Arntl^{fl}$ and $Arntl^{\Delta Rorgt}$ mice revealed a profound reduction in the
expression of reactivity genes in the $Arntl^{\Delta Rorgt}$ intestinal epithelium;
notably, *Reg3b*, *Reg3g*, *Muc3* and *Muc13* were consistently reduced in
*Arntl*-deficient mice (Fig. 2c). Furthermore, $Arntl^{\Delta Rorgt}$ mice displayed
altered diurnal patterns of Proteobacteria and Bacteroidetes (Fig. 2d,
Extended Data Fig. 3j). To investigate whether disruption of ILC3-
intrinsic ARNTL affected enteric defence, we tested how $Arntl^{\Delta Rorgt}$
mice responded to intestinal infection. To this end, we bred $Arntl^{\Delta Rorgt}$
mice to $Rag1^{-/-}$ mice to exclude putative T cell effects (Extended Data
Fig. 3g–i). $Rag1^{-/-}Arntl^{\Delta Rorgt}$ mice were infected with the attaching
and effacing bacteria *Citrobacter rodentium*[2]. When compared to their

wild-type littermate controls, $Rag1^{-/-}Arntl^{\Delta Rorgt}$ mice had marked gut
inflammation, fewer IL-22-producing ILC3s, increased *C. rodentium*
infection and bacterial translocation, reduced expression of epithelial
reactivity genes, increased weight loss and reduced survival (Fig. 2e–j,
Extended Data Fig. 5a–j). These results indicate that cell-intrinsic circa-
dian signals selectively control intestinal ILC3s and shape gut epithelial
reactivity, microbial communities and enteric defence. Previous studies
indicated that ILC3s regulate host lipid metabolism[8]. When compared
to their wild-type littermate controls, the epithelium of $Arntl^{\Delta Rorgt}$ mice
revealed a marked increase in mRNA that codes for key lipid epithelial
transporters, including *Fabp1*, *Fabp2*, *Scd1*, *Cd36* and *Apoe* (Fig. 2k).
Accordingly, these changes were associated with increased gonadal and
subcutaneous accumulation of fat in $Arntl^{\Delta Rorgt}$ mice when compared
to their wild-type littermate controls (Fig. 2l, Extended Data Fig. 5k–n).
Thus, ILC3-intrinsic circadian signals shape epithelial lipid transport
and body fat composition.

To further investigate how cell-intrinsic *Arntl* controls intestinal
ILC3 homeostasis, initially we studied the diurnal oscillations of the
ILC3 clock machinery. When compared to their wild-type littermate
controls, $Arntl^{\Delta Rorgt}$ ILC3s displayed a disrupted diurnal pattern of
activator and repressor circadian genes (Fig. 3a). Sequentially, we used
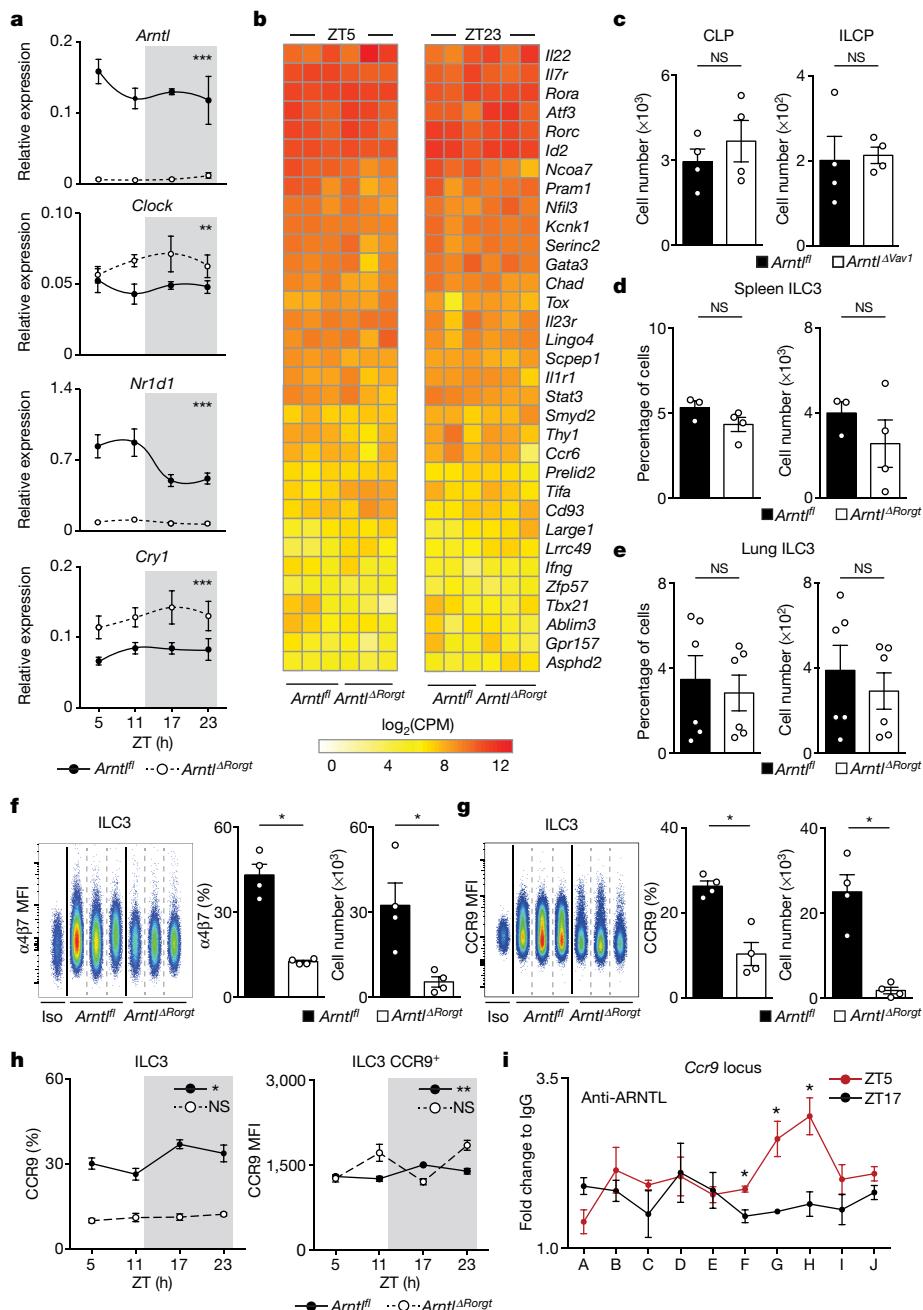
**Fig. 3 | ILC3-intrinsic circadian signals regulate an enteric receptor postcode. a**, Relative expression of circadian genes in enteric ILC3s from $Arntl^{fl}$ and $Arntl^{\Delta Rorgt}$ mice; $n = 3$. **b**, RNA sequencing (RNA-seq) analysis of gut ILC3s from $Arntl^{fl}$ and $Arntl^{\Delta Rorgt}$ mice at ZT5 and ZT23; $n = 3$. **c**, Numbers of CLPs and ILCPs in $Arntl^{fl}$ and $Arntl^{\Delta Vav1}$ mice; $n = 4$. **d**, **e**, ILC3s in spleen (**d**, $n = 3$) and lung (**e**, $n = 6$) of $Arntl^{fl}$ and $Arntl^{\Delta Rorgt}$ mice. **f**, **g**, Expression of α4β7 (**f**) and CCR9 (**g**) by gut ILC3s in $Arntl^{fl}$ and $Arntl^{\Delta Rorgt}$ mice; $n = 4$. **h**, Circadian variation in expression of CCR9 by intestinal ILC3s in $Arntl^{fl}$ and $Arntl^{\Delta Rorgt}$ mice; $n = 4$. **i**, ChIP analysis of binding of ARNTL to the $Ccr9$ locus in enteric ILC3s; $n = 3$. A–J denote putative ARNTL DNA-binding sites. Data are representative of three independent experiments. $n$ represents biologically independent animals (**a**, **c**–**h**) or samples (**b**, **i**). **a**, **h**, White and grey represent light and dark periods, respectively. Data shown as mean ± s.e.m. **a**, Two-way ANOVA; **c**–**g**, two-tailed Mann–Whitney $U$ test; **h**, cosinor analysis; **i**, two-tailed unpaired Student's $t$-test. *$P < 0.05$; **$P < 0.01$; ***$P < 0.001$; NS, not significant.

genome-wide transcriptional profiling of $Arntl$-sufficient and -deficient ILC3s to interrogate the effect of a deregulated circadian machinery. Diurnal analysis of the genetic signature associated with ILC3 identity[1] demonstrated that the vast majority of those genes were unperturbed in $Arntl$-deficient ILC3s, suggesting that ARNTL is dispensable to ILC3 lineage commitment (Fig. 3b, Extended Data Fig. 6a–c). To test this hypothesis, we first studied the effect of ablation of $Arntl$ in ILC3 progenitors. $Arntl^{\Delta Vav1}$ mice had unperturbed numbers of common lymphoid progenitors (CLPs) and innate lymphoid cell progenitors (ILCPs; Fig. 3c, Extended Data Fig. 6d). Sequentially, we analysed the effects of

$Arntl$ ablation in ILC3s in other organs. Compared to their littermate controls, $Arntl^{\Delta Rorgt}$ mice had normal numbers of ILC3s in the spleen, lungs and blood, in contrast to their pronounced reduction in the intestine (Figs. 2a, 3d, e, Extended Data Fig. 6e). Notably, enteric $Arntl^{\Delta Rorgt}$ ILC3s showed unperturbed proliferation and apoptosis-related genetic signatures (Extended Data Fig. 6b, c), suggesting that $Arntl^{\Delta Rorgt}$ ILC3s may show altered migration to the intestinal mucosa[9]. When compared to their wild-type littermate controls, ILC3s in $Arntl^{\Delta Rorgt}$ mice showed a marked reduction in gut postcode molecules—which are essential receptors for intestinal lamina propria homing—and accumulated in
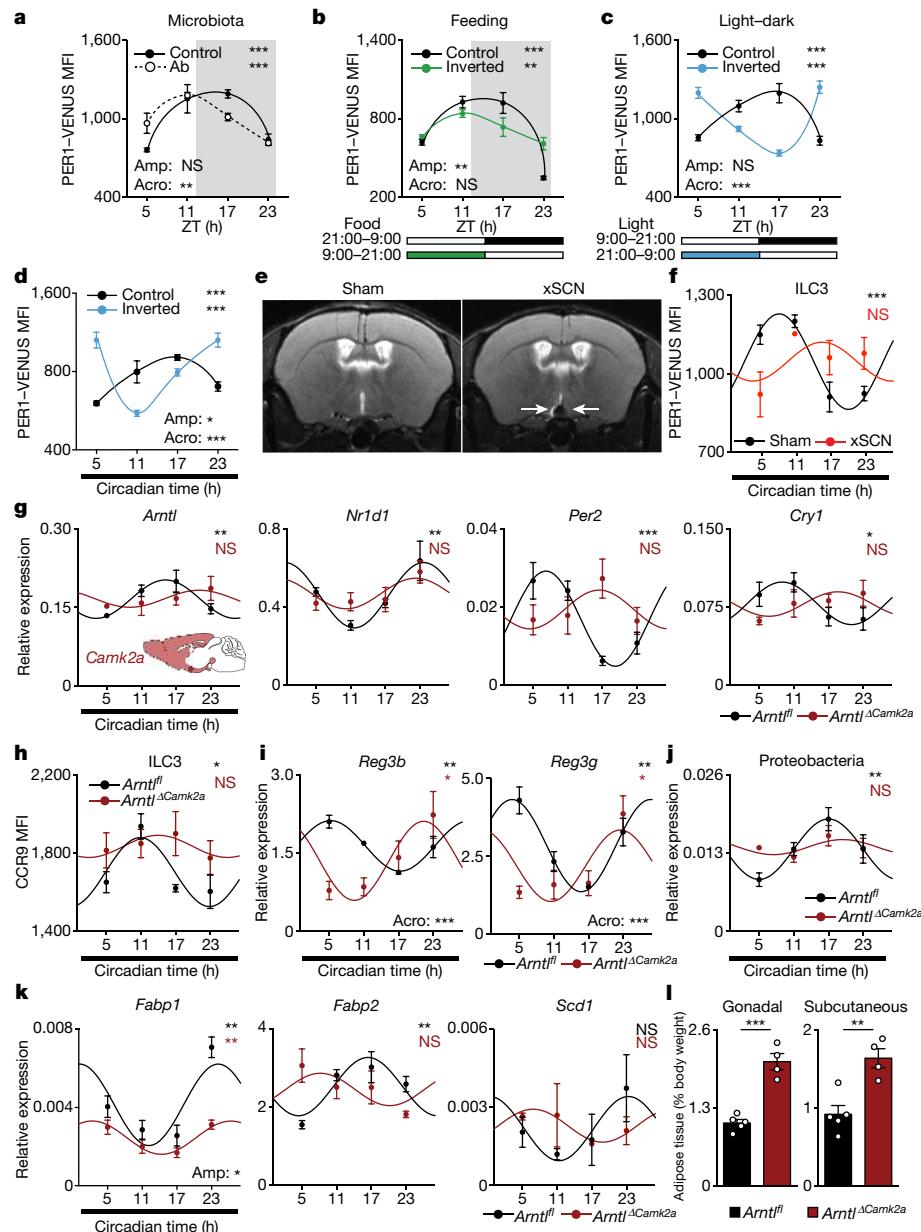
**Fig. 4 | Light-entrained and brain-tuned cues shape intestinal ILC3s.**
**a–d**, PER1–VENUS MFI in gut ILC3s from mice treated with or without
antibiotics (Ab) (**a**; $n = 3$); with restricted or inverted feeding (**b**; $n = 3$);
with opposing light–dark cycles (**c**; $n = 3$); and with opposing light–dark
cycles followed by constant darkness (**d**; $n = 3$). **e**, Magnetic resonance
imaging of sham- and SCN-ablated (xSCN) mice; $n = 11$. White arrows
indicate location of lesion. **f**, PER1–VENUS MFI in enteric ILC3s from
sham- or SCN-ablated mice; $n = 3$. **g**, Expression of circadian genes
in enteric ILC3s from $Arntl^{fl}$ and $Arntl^{\Delta Camk2a}$ mice; $n = 3$. **h**, CCR9
expression in gut ILC3s from $Arntl^{fl}$ and $Arntl^{\Delta Camk2a}$ mice; $n = 3$.
**i**, Expression of epithelial reactivity genes in the small intestine from

$Arntl^{fl}$ and $Arntl^{\Delta Camk2a}$ mice; $n = 3$. **j**, qPCR analysis of Proteobacteria
in stools from $Arntl^{fl}$ and $Arntl^{\Delta Camk2a}$ mice; $n = 4$. **k**, Expression of
lipid transporter genes in the epithelium of the small intestine in $Arntl^{fl}$
and $Arntl^{\Delta Camk2a}$ mice; $n = 3$. **l**, Gonadal and subcutaneous adipose
tissue in $Arntl^{fl}$ mice ($n = 5$) and $Arntl^{\Delta Camk2a}$ mice ($n = 4$). **a**, **b**, White
and grey represent light and dark periods, respectively. Data shown
as mean ± s.e.m. $n$ represents biologically independent animals.
**a–d**, **f–k**, Cosinor analysis; **f–k**, cosine fitted curves; amplitude (Amp) and
acrophase (Acro) were extracted from the cosinor model. **l**, Two-tailed
unpaired Student's $t$-test. *$P < 0.05$; **$P < 0.01$; ***$P < 0.001$; NS, not
significant.

mesenteric lymph nodes[9] (Extended Data Fig. 6f). Notably, the expression of the integrin and chemokine receptors CCR9, α4β7 and CXCR4 was selectively and hierarchically reduced in $Arntl^{\Delta Rorgt}$ ILC3s (Fig. 3f–h, Extended Data Fig. 6g–m). To investigate whether ARNTL could directly regulate expression of $Ccr9$, we performed chromatin immunoprecipitation (ChIP). Binding of ARNTL to the $Ccr9$ locus in ILC3s followed a diurnal pattern, with increased binding at Zeitgeber time (ZT) 5 (Fig. 3i). Thus, ARNTL can contribute directly to the expression of $Ccr9$ in ILC3s, although additional factors may also regulate this gene. In conclusion, while a fully operational ILC3-intrinsic circadian machinery is not required for lineage commitment and development

of ILC3s, cell-intrinsic clock signals are required for a functional ILC3 gut receptor postcode.

Circadian rhythms allow organisms to adapt to extrinsic environmental changes. Microbial cues can alter the rhythms of intestinal cells[10,11], and feeding regimens are major circadian entraining cues for peripheral organs, such as the liver[12]. In order to define the environmental cues that entrain circadian oscillations of ILC3, we initially investigated whether microbial signals affect the oscillations of ILC3s. Treatment of $Per1^{Venus}$ reporter mice with antibiotics did not alter the amplitude of circadian oscillations, but did induce a minute shift in the acrophase (timing of the peak of the cycle; Fig. 4a). We then

tested whether feeding regimens, which are major entraining cues of oscillations in the liver, pancreas, kidney, and heart[12], could alter ILC3 rhythms. To this end, we restricted food access to a 12-h interval and compared $Per1^{Venus}$ oscillations to those observed in mice with inverted feeding regimens[12]. Inverted feeding had a small effect on the amplitude of ILC3 oscillations but did not invert the acrophase of ILC3s (Fig. 4b, Extended Data Fig. 7a), in contrast to the full inversion of the acrophase of hepatocytes[12] (Extended Data Fig. 7b). As these local intestinal cues could not invert the acrophase of ILC3s, we hypothesize that light–dark cycles are major regulators of enteric ILC3 oscillations[6]. To test this hypothesis, we placed $Per1^{Venus}$ mice in light-tight cabinets on two opposing 12-h light–dark cycles. Inversion of light–dark cycles had a profound effect on the circadian oscillations of ILC3s (Fig. 4c). Notably, and in contrast to microbiota and feeding regimens, light cycles fully inverted the acrophase of $Per1^{Venus}$ oscillations in ILC3s (Fig. 4c, Extended Data Fig. 7c). Furthermore, light–dark cycles entrained ILC3 oscillations, as revealed by their maintenance upon removal of light (constant darkness; Fig. 4d, Extended Data Fig. 7d), confirming that light is a major environmental entraining signal for ILC3 intrinsic oscillations. Together, these data indicate that ILC3s integrate systemic and local cues hierarchically; while microbiota and feeding regimens locally adjust the ILC3 clock, light–dark cycles are major entraining cues of ILC3s, fully setting and entraining their intrinsic oscillatory clock.

The suprachiasmatic nuclei (SCN) in the hypothalamus are main integrators of light signals[6], suggesting that brain cues may regulate ILC3s. To assess the influence of the master circadian pacemaker on ILC3s, while excluding confounding light-induced, SCN-independent effects[13,14], we performed SCN ablation by electrolytic lesion in $Per1^{Venus}$ mice using stereotaxic brain surgery[15]. Strikingly, whereas sham-operated mice displayed circadian $Per1^{Venus}$ oscillations in ILC3s, ILC3s in SCN-ablated mice lost the circadian rhythmicity of $Per1^{Venus}$ and other circadian genes (Fig. 4e, f, Extended Data Fig. 8a–d). Because electrolytic lesions of the SCN may cause scission of afferent and efferent fibres in the SCN, we further confirmed that brain SCN-derived cues control ILC3s by genetic ablation of $Arntl$ in the SCN[14]. $Arntl^{fl}$ mice were bred to $Camk2a^{Cre}$ mice to allow forebrain- and SCN-specific deletion of $Arntl$ ($Arntl^{\Delta Camk2a}$)[14]. When compared to their control counterparts, ILC3s from $Arntl^{\Delta Camk2a}$ mice showed severe arrhythmicity of circadian regulatory genes and of the enteric postcode molecule CCR9 (Fig. 4g, h, Extended Data Fig. 9a–f). In addition, $Arntl^{\Delta Camk2a}$ mice showed alterations in epithelial reactivity genes and microbial communities, particularly Proteobacteria and Bacteroidetes (Fig. 4i, j, Extended Data Fig. 9g–i). Finally, the intestinal epithelium of $Arntl^{\Delta Camk2a}$ mice showed disrupted circadian expression of lipid epithelial transporters, and these changes were associated with increased gonadal and subcutaneous fat accumulation (Fig. 4k, l). Together, these data indicate that light-entrained and brain-tuned circuits regulate enteric ILC3s, controlling microbial communities, lipid metabolism and body composition.

Deciphering the mechanisms by which neuroimmune circuits operate to integrate extrinsic and systemic signals is essential for understanding tissue and organ homeostasis. We found that light cues are major extrinsic entraining cues of ILC3 circadian rhythms, and surgically or genetically induced deregulation of brain rhythmicity resulted in altered ILC3 regulation. In turn, the ILC3-intrinsic circadian machinery controlled the gut receptor postcode of ILC3s, shaping enteric ILC3s and host homeostasis.

Our data reveal that ILC3s display diurnal oscillations that are genetically encoded, cell-autonomous and entrained by light cues. While microbiota and feeding regimens could locally induce small adjustments to ILC3 oscillations, light–dark cycles were major entraining cues of the ILC3 circadian clock. Whether the effects of photonic signals on ILC3s are immediate or rely on other peripheral clocks remains to be elucidated[16,17]. Nevertheless, cell-intrinsic ablation of important endocrine and peripheral neural signals in ILC3s did not affect gut

ILC3 numbers (Extended Data Fig. 10a-i). Our work indicates that ILC3s integrate local and systemic entraining cues in a distinct hierarchical manner, establishing an organismal circuitry that is an essential link between the extrinsic environment, enteric ILC3s, gut defence, lipid metabolism and host homeostasis (Extended Data Fig. 10j).

Previous studies demonstrated that ILCs integrate tissue microenvironmental signals, including cytokines, micronutrients and neuroregulators[3,4,18,19]. Here we show that ILC3s have a cell-intrinsic circadian clock that integrates extrinsic light-entrained and brain-tuned signals. Coupling light cues to ILC3 circadian regulation may have ensured efficient and integrated multi-system anticipatory responses to environmental changes. Notably, the regulation of ILC3 activity by systemic circadian circuits may have evolved to maximize metabolic homeostasis, gut defence and efficient symbiosis with commensal organisms that have been evolutionary partners of mammals. Finally, our current data may also contribute to a better understanding of how circadian disruptions in humans are associated with metabolic diseases, bowel inflammatory conditions and cancer[20].

## Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-019-1579-3.

1. Vivier, E. et al. Innate lymphoid cells: 10 years on. *Cell* **174**, 1054–1066 (2018).
2. Ibiza, S. et al. Glial-cell-derived neuroregulators control type 3 innate lymphoid cells and gut defence. *Nature* **535**, 440–443 (2016).
3. Veiga-Fernandes, H. & Artis, D. Neuronal-immune system cross-talk in homeostasis. *Science* **359**, 1465–1466 (2018).
4. Godinho-Silva, C., Cardoso, F. & Veiga-Fernandes, H. Neuro-immune cell units: a new paradigm in physiology. *Annu. Rev. Immunol.* **37**, 19–46 (2019).
5. Veiga-Fernandes, H. & Pachnis, V. Neuroimmune regulation during intestinal development and homeostasis. *Nat. Immunol.* **18**, 116–122 (2017).
6. Takahashi, J. S. Transcriptional architecture of the mammalian circadian clock. *Nat. Rev. Genet.* **18**, 164–179 (2017).
7. Scheiermann, C., Gibbs, J., Ince, L. & Loudon, A. Clocking in to immunity. *Nat. Rev. Immunol.* **18**, 423–437 (2018).
8. Mao, K. et al. Innate and adaptive lymphocytes sequentially shape the gut microbiota and lipid metabolism. *Nature* **554**, 255–259 (2018).
9. Mackley, E. C. et al. CCR7-dependent trafficking of RORγ⁺ ILCs creates a unique microenvironment within mucosal draining lymph nodes. *Nat. Commun.* **6**, 5862 (2015).
10. Thaiss, C. A. et al. Microbiota diurnal rhythmicity programs host transcriptome oscillations. *Cell* **167**, 1495–1510 (2016).
11. Wang, Y. et al. The intestinal microbiota regulates body composition through NFIL3 and the circadian clock. *Science* **357**, 912–916 (2017).
12. Damiola, F. et al. Restricted feeding uncouples circadian oscillators in peripheral tissues from the central pacemaker in the suprachiasmatic nucleus. *Genes Dev.* **14**, 2950–2961 (2000).
13. Husse, J., Leliavski, A., Tsang, A. H., Oster, H. & Eichele, G. The light-dark cycle controls peripheral rhythmicity in mice with a genetically ablated suprachiasmatic nucleus clock. *FASEB J.* **28**, 4950–4960 (2014).
14. Izumo, M. et al. Differential effects of light and feeding on circadian organization of peripheral clocks in a forebrain *Bmal1* mutant. *eLife* **3**, e04617 (2014).
15. Sujino, M. et al. Suprachiasmatic nucleus grafts restore circadian behavioral rhythms of genetically arrhythmic mice. *Curr. Biol.* **13**, 664–668 (2003).
16. Koronowski, K. B. et al. Defining the independence of the liver circadian clock. *Cell* **177**, 1448–1462 (2019).
17. Welz, P. S. et al. BMAL1-driven tissue clocks respond independently to light to maintain homeostasis. *Cell* **177**, 1436–1447 (2019).
18. van de Pavert, S. A. et al. Maternal retinoids control type 3 innate lymphoid cells and set the offspring immunity. *Nature* **508**, 123–127 (2014).
19. Cardoso, V. et al. Neuronal regulation of type 2 innate lymphoid cells via neuromedin U. *Nature* **549**, 277–281 (2017).
20. Masri, S. & Sassone-Corsi, P. The emerging link between cancer, metabolism, and circadian rhythms. *Nat. Med.* **24**, 1795–1803 (2018).

## METHODS

**Mice.** Nod scid gamma (NSG) mice were purchased from Jackson Laboratories. C57BL/6J Ly5.1 mice were purchased from Jackson Laboratories and bred with C57BL/6J mice to obtain C57BL/6 Ly5.1/Ly5.2 (CD45.1/CD45.2). Mouse lines used were: $Rag1^{-/-}$ (ref. [21]), $Rag2^{-/-}Il2rg^{-/-}$ (ref. [22,23]), $Vav1^{Cre}$ (ref. [24]), $Rorgt^{Cre}$ (ref. [25]), $Camk2a^{Cre}$ (ref. [26]), $Il7ra^{Cre}$ (ref. [27]), $Per1^{Venus}$ (ref. [28]), $Ret^{GFP}$ (ref. [29]), $Rosa26^{RFP}$ (ref. [30]), $Nr1d1^{-/-}$ (ref. [31]), $Arntl^{fl}$ (ref. [32]), $Nr3c1^{fl}$ (ref. [33]) and $Adrb2^{fl}$ (ref. [34]). All mouse lines were on a full C57BL/6J background. All lines were bred and maintained at Champalimaud Centre for the Unknown (CCU) animal facility under specific pathogen-free conditions. Male and female mice were used at 8–14 weeks old, unless stated otherwise. Sex- and age-matched mice were used for analysis of small intestine epithelium lipid transporters and quantification of white adipose tissue. Mice were maintained in 12-h light–dark cycles, with ad libitum access to food and water, if not specified otherwise. For light inversion experiments mice were housed in ventilated, light-tight cabinets on defined 12-h light–dark cycles (Ternox). $Camk2a^{Cre}Arntl^{fl}$ ($Arntl^{\Delta Camk2a}$) mice and wild-type littermate controls were maintained in constant darkness as previously described[14]. Mice were systematically compared with co-housed littermate controls unless stated otherwise. Power analysis was performed to estimate the number of experimental mice required. All animal experiments were approved by national and local institutional review boards (IRBs), Direção Geral de Veterinária and CCU ethical committees. Randomization and blinding were not used unless stated otherwise.

**Cell isolation.** Isolation of small intestine and colonic lamina propria cells was as previously described[2]. In brief, intestines and colons were thoroughly rinsed with cold PBS1×, Peyer patches were removed from the small intestine, and intestines and colons were cut into 1-cm pieces and shaken for 30 min in PBS containing 2% FBS, 1% HEPES and 5 mM EDTA to remove intraepithelial and epithelial cells. Intestines and colons were then digested with collagenase D (0.5 mg/ml; Roche) and DNase I (20 U/ml; Roche) in complete RPMI for 30 min at 37 °C, under gentle agitation. Cells were passed through a 100-μm cell strainer and purified by centrifugation for 30 min at 2,400 rpm in a 40/80 Percoll (GE Healthcare) gradient. Lungs were finely minced and digested in complete RPMI supplemented with collagenase D (0.1 mg/ml; Roche) and DNase I (20 U/ml; Roche) for 1 h at 37 °C under gentle agitation. Cells were passed through a 100-μm cell strainer and purified by centrifugation for 30 min at 2,400 rpm in a 40/80 Percoll (GE Healthcare) gradient. Spleen and mesenteric lymph node cell suspensions were obtained using 70-μm strainers. Bone marrow cells were collected by either flushing or crushing bones and filtered using 70-μm strainers. Erythrocytes from small intestine, colon, lung, spleen and bone marrow preparations were lysed with RBC lysis buffer (eBioscience). Leukocytes from blood were isolated by treatment with Ficoll (GE Healthcare).

**Flow cytometry analysis and cell sorting.** For cytokine analysis ex vivo, cells were incubated with PMA (phorbol 12-myristate 13-acetate; 50 ng/ml) and ionomycin (500 ng/ml) (Sigma-Aldrich) in the presence of brefeldin A (eBioscience) for 4 h before intracellular staining. Intracellular staining for cytokines and transcription factors analysis was performed using IC fixation and Staining Buffer Set (eBioscience). Cell sorting was performed using FACSFusion (BD Biosciences). Sorted populations were >95% pure. Flow cytometry analysis was performed on LSRFortessa X-20 (BD Biosciences). Data were analysed using FlowJo 8.8.7 software (Tree Star). Cell populations were gated in live cells, both for sorting and flow cytometry analysis.

**Cell populations.** Cell populations were defined as: bone marrow (BM) common lymphoid progenitor (CLP): Lin$^-$CD127$^+$Flt3$^+$Sca1$^{int}$c-Kit$^{int}$; BM innate lymphoid cell progenitor (ILCP): Lin$^-$CD127$^+$Flt3$^-$CD25$^-$c-Kit$^+$α4β7$^{high}$; BM ILC2 progenitor (ILC2P): Lin$^-$CD127$^+$Flt3$^-$Sca1$^+$CD25$^+$; small intestine (SI) NK: CD45$^+$Lin$^-$NK1.1$^+$NKp46$^+$CD27$^+$CD49b$^+$CD127$^-$ EOMES$^+$ or CD45$^+$Lin$^-$NK1.1$^+$NKp46$^+$CD27$^+$CD49b$^+$CD127$^-$; small intestine ILC1: CD45$^+$Lin$^-$NK1.1$^+$NKp46$^+$CD27$^+$CD49b$^-$CD127$^-$Tbet$^+$ or CD45$^+$Lin$^-$NK1.1$^+$NKp46$^+$CD27$^+$CD49b$^-$CD127$^+$; small intestine ILC2: CD45$^+$Lin$^-$Thy1.2$^+$KLRG1$^+$GATA3$^+$ or CD45$^+$Lin$^-$Thy1.2$^+$KLRG1$^+$Sca-1$^+$CD25$^+$; lamina propria, spleen, mesenteric lymph node and lung ILC3: CD45$^+$Lin$^-$Thy1.2$^{high}$RORγt$^+$ or CD45$^+$Lin$^-$Thy1.2$^{high}$KLRG1$^-$; ILC3-IL-17$^+$: CD45$^+$Lin$^-$Thy1.2$^{high}$RORγt$^+$IL-17$^+$; ILC3-IL-22$^+$: CD45$^+$Lin$^-$Thy1.2$^{high}$RORγt$^+$IL-22$^+$; for ILC3 subsets additional markers were used: ILC3-NCR$^-$CD4$^-$: NKp46$^-$CD4$^-$; ILC3-LTi CD4$^+$: NKp46$^-$CD4$^+$; ILC3-CCR6$^-$NCR$^-$: CCR6$^-$NKp46$^-$; ILC3-LTi-like: CCR6$^+$NKp46$^-$; ILC3-NCR$^+$: NKp46$^+$; SI Th17 cells: CD45$^+$Lin$^+$Thy1.2$^+$CD4$^+$RORγt$^+$; colon Tregs: CD45$^+$CD3$^+$Thy1.2$^+$CD4$^+$CD25$^+$FOXP3$^+$; colon Tregs RORγt$^+$: CD45$^+$CD3$^+$Thy1.2$^+$CD4$^+$CD25$^+$FOXP3$^+$RORγt$^+$. The lineage cocktail for BM, lung, small intestine lamina propria, spleen and mesenteric lymph nodes included CD3ε, CD8α, CD19, B220, CD11c, CD11b, Ter119, Gr1, TCRβ, TCRγδ and NK1.1. For NK and ILC1 staining in the small intestine, NK1.1 and CD11b were not added to the lineage cocktail.

**Antibody list.** Cell suspensions were stained with: anti-CD45 (30-F11); anti-CD45.1 (A20); anti-CD45.2 (104); anti-CD11c (N418); anti-CD11b (Mi/70);

anti-CD127 (IL7Rα; A7R34); anti-CD27(LG.7F9); anti-CD8α (53-6.7); anti-CD19 (eBio1D3); anti-CXCR4(L276F12); anti-NK1.1 (PK136); anti-CD3ε (eBio500A2); anti-TER119 (TER-119); anti-Gr1 (RB6-8C5); anti-CD4 (RM4-5); anti-CD25 (PC61); anti-CD117 (c-Kit; 2B8); anti-CD90.2 (Thy1.2; 53-2.1); anti-TCRβ (H57-595); anti-TCRγδ (GL3); anti-B220 (RA3-6B2); anti-KLRG1 (2F1/KLRG1); anti-Ly-6A/E (Sca1; D7); anti-CCR9 (CW-1.2); anti-IL-17 (TC11-18H10.1); anti-rat IgG1k isotype control (RTK2071); anti-streptavidin fluorochrome conjugates from Biolegend; anti-α4β7 (DATK32); anti-Flt3 (A2F10); anti-NKp46 (29A1.4); anti-CD49b (DX5); anti-Ki67 (SolA15); anti-rat IgG2ak isotype control (eBR2a); anti-IL-22 (1H8PWSR); anti-rat IgG1k isotype control (eBRG1); anti-EOMES (Dan11mag); anti-Tbet (eBio4B10); anti-FOPX3 (FJK-16s); anti-GATA3 (TWAJ); anti-CD16/CD32 (93); 7AAD viability dye from eBiosciences; anti-CD196 (CCR6; 140706) from BD Biosciences; anti-RORγt (Q31-378) and anti-mouse IgG2ak isotype control (G155-178) from BD Pharmingen. LIVE/DEAD Fixable Aqua Dead Cell Stain Kit was purchased from Invitrogen.

**Bone marrow transplantation.** Bone marrow CD3$^-$ cells were FACS sorted from $Arntl^{fl}$, $Vav1^{Cre}Arntl^{fl}$, $Rag1^{-/-}Arntl^{fl}$, $Rag1^{-/-}Rorgt^{Cre}Arntl^{fl}$, $Nr1d1^{+/+}$, $Nr1d1^{-/-}$ and C57BL/6 Ly5.1/Ly5.2 mice. Sorted cells ($2 \times 10^5$) from $Arntl$- or $Nr1d1$-deficient and -competent wild-type littermate controls were intravenously injected in direct competition with a third-party wild-type competitor (CD45.1/CD45.2), in a 1:1 ratio, into non-lethally irradiated NSG (150cGy) or $Rag2^{-/-}Il2rg^{-/-}$ (500cGy) mice (CD45.1). Recipients were analysed 8 weeks after transplantation.

**Quantitative RT–PCR.** RNA from sorted cells was extracted using RNeasy micro kit (Qiagen) according to the manufacturer's protocol. Liver, small intestine (ileum) and colon epithelium was collected for RNA extraction using Trizol (Invitrogen) and zirconia/silica beads (BioSpec) in a bead beater (MIDSCI). RNA concentration was determined using Nanodrop Spectrophotometer (Nanodrop Technologies). For TaqMan assays (Applied Biosystems) RNA was retro-transcribed using a High Capacity RNA-to-cDNA Kit (Applied Biosystems), followed by a pre-amplification PCR using TaqMan PreAmp Master Mix (Applied Biosystems). TaqMan Gene Expression Master Mix (Applied Biosystems) was used in real-time PCR. Real-time PCR analysis was performed using StepOne and QuantStudio 5 Real-Time PCR systems (Applied Biosystems). $Hprt$, $Gapdh$ and $Eef1a1$ were used as housekeeping genes. When multiple endogenous controls were used, these were treated as a single population and the reference value calculated by arithmetic mean of their CT values. The mRNA analysis was performed as previously described[35]. In brief, we used the comparative $C_T$ method ($2^{-\Delta CT}$), in which $\Delta C_{T(gene of interest)} = C_{T(gene of interest)} - C_{T(housekeeping reference value)}$. When fold change comparison between samples was required, the comparative $\Delta C_T$ method ($2^{-\Delta\Delta CT}$) was applied.

**TaqMan gene expression assays.** TaqMan Gene Expression Assays (Applied Biosystems) were the following: $Hprt$ Mm00446968_m1; $Gapdh$ Mm99999915_g1; $Eef1a1$ Mm01973893_g1; $Arntl$ Mm00500223_m1; $Clock$ Mm00455950_m1; $Nr1d1$ Mm00520708_m1; $Nr1d2$ Mm01310356_g1; $Per1$ Mm00501813_m1; $Per2$ 00478113_m1; $Cry1$ Mm00500223_m1; $Cry2$ Mm01331539_m1; $Runx1$ Mm01213404_m1; $Tox$ Mm00455231_m1; $Rorgt$ Mm01261022_m1; $Ahr$ Mm00478932_m1; $Rora$ Mm01173766_m1; $Ccr9$ Mm02528165_s1; Reg3a Mm01181787_m1; $Reg3b$ Mm00440616_g1; $Reg3g$ Mm00441127_m1; $Muc1$ Mm00449604_m1; $Muc2$ Mm01276696_m1; $Muc3$ Mm01207064_m1; $Muc13$ Mm00495397_m1; $S100a8$ Mm01276046_m1; $S100a9$ Mm00656925_m1; $Epcam$ Mm00493214_m1; $Apoe$ Mm01307193_g1; $Cd36$ Mm01307193_g1; $Fabp1$ Mm00444340_m1; $Fabp2$ Mm00433188_m1; and $Scd1$ Mm00772290_m1.

**Quantitative PCR analysis of bacteria in stools at the phylum level.** DNA from faecal pellets of female mice was isolated with ZR Fecal DNA MicroPrep (Zymo Research). Quantification of bacteria was determined from standard curves established by qPCR as previously described[2]. qPCRs were performed with NZY qPCR Green Master Mix (Nzytech) and different primer sets using a QuantStudio 5 Real-Time PCR System (Applied Biosystems) thermocycler. Samples were normalized to 16S rDNA and reported according to the $2^{-\Delta CT}$ method. Primer sequences were: 16S rDNA, F-ACTCCTACGGGAGGCAGCAGT and R-ATTACCGCGGCTGCTGGC; Bacteroidetes, F-GAGAGGAAGGTCCCCCAC and R-CGCTACTTGGCTGGTTCAG; Proteobacteria, F-GGTTCTGAGAGGA GGTCCC and R-GCTGGCTCCCGTAGGAGT; Firmicutes, F-GGAGCATGTG GTTTAATTCGAAGCA and R-AGCTGACGACAACCATGCAC.

***C. rodentium* infection.** Infection with *C. rodentium* ICC180 (derived from DBS100 strain)[36] was performed at ZT6 by gavage inoculation with $10^9$ colony-forming units (CFUs)[36,37]. Acquisition and quantification of luciferase signal was performed in an IVIS Lumina III System (Perkin Elmer). Throughout infection, weight loss, diarrhoea and bloody stools were monitored daily.

**CFU measurement.** Bacterial translocation was determined in the spleen, liver, and mesenteric lymph nodes, taking in account total bacteria and luciferase-positive *C. rodentium*. Organs were removed, weighed and brought into suspension. Bacterial CFUs from organ samples were determined via serial dilutions on Luria broth (LB) agar (Invitrogen) and MacConkey agar (Sigma-Aldrich). Colonies were counted after 2 days of culture at 37 °C. Luciferase-positive *C. rodentium* was

quantified on MacConkey agar plates using an IVIS Lumina III System (Perkin Elmer). CFUs were determined per volume (ml) for each organ.

**Antibiotic and dexamethasone treatment.** Pregnant females and newborn mice were treated with streptomycin (5 g/l), ampicillin (1 g/l) and colistin (1 g/l) (Sigma-Aldrich) in drinking water with 3% sucrose. Control mice were given 3% sucrose in drinking water as previously described[38]. Dexamethasone 21-phosphate disodium salt (200 μg) (Sigma) or PBS was injected intraperitoneally at ZT0. After 4, 8, 12 and 23 h (ZT 4, 8, 12 and 23) mice were killed and analysed.

**ChIP assay.** Enteric ILC3s from adult C57BL/6J mice were isolated by flow cytometry. Cells were fixed, cross-linked and lysed, and chromosomal DNA–protein complexes were sonicated to generate DNA fragments ranging from 200 to 400 base pairs as previously described[2]. DNA–protein complexes were immunoprecipitated using LowCell# ChIP kit (Diagenode), with 1 μg of antibody against ARNTL (Abcam) and IgG isotype control (Abcam). Immunoprecipitates were uncrosslinked and analysed by qPCR using primer pairs flanking ARNTL putative sites (E-boxes) in the *Ccr9* locus (determined by computational analysis using TFBS tools and Jaspar 2018). Results were normalized to input intensity and control IgG. Primer sequences were: A: F-CATTTCATAGCTTAGGCTGGCATGG; R-CTAGCTAACTGGTCTCAAAGTCCTC; B: F-GCCTCCCTTGTACTACCTG AAGC; R-TCCCAACACCAGGCCGAGTA; C: F-AGGGTCAATTTCTT AGGGCGACA; R-GCCAAGTGTTCGGTCCCAC; D: F-TCTGGCTTCT CACCATGACCACT; R-TCTAAGGCGTCACCACTGTTCTC, E: F-TTTGG GGAATCATCTTACAGC AGAG; R-ATTCATCCTGGCCCTTTCCTTCTTA; F: F-GCTCCACCTCATAGTTGTCTGG; R-CCATGAGCACGTGGAGAGAAAG; G: F-GGTCGAATACCGCGTGGGTT; R-CCCGGTAGAGGCTGCAAGAAA; H: F-AGGCAAATCTGGGCCTATCC; R-GGCCCAGTACAGAGGGGTCT; I: F-GGCTCAGGCTAGCAGGTCTC; R-TGTTTGGCCAGGCATCCTCCA; J: F-ACTCAGAGGTGCTGTGACTCC; R-AGCTTTAGGACCACAATGGGCA.

**Food restriction (inverted feeding).** *Per1^Venus* mice fed during the night received food from 21:00 to 9:00 (control group), whereas mice fed during the day had access to food from 9:00 to 21:00 (inverted group). Food restriction was performed during nine consecutive days as previously described[12]. For food restriction in constant darkness, *Per1^Venus* mice were housed in constant darkness with ad libitum access to food and water for 2 weeks. Then, access to food was restricted to the subjective day or night, for 12 days, in constant darkness.

**Inverted light–dark cycles.** To induce changes in light regime, *Per1^Venus* mice were placed in ventilated, light-tight cabinets on a 12-h light–dark cycle (Ternox). After acclimation, light cycles were changed for mice in the inverted group for 3 weeks to completely establish an inverse light cycle, while they remained the same for mice in the control group, as previously described[39]. For inverted light–dark cycle experiments followed by constant darkness, after establishing an inverse light–dark cycle, mice were transferred into constant darkness for 3 weeks.

**SCN lesions.** Bilateral ablation of the SCN was performed in 9–12-week-old *Per1^Venus* males by electrolytic lesion using stereotaxic brain surgery, as described previously[15]. Mice were kept under deep anaesthesia using a mixture of isoflurane and oxygen (1–3% isoflurane at 1 l/min). Surgeries were performed using a stereotaxic device (Kopf). After identification of the bregma, a hole was drilled through which the lesion electrode was inserted into the brain. Electrodes were made by isolating a 0.25-mm stainless steel insect pin with a heat shrink polyester tubing, except for 0.2 mm at the tip. The electrode tip was aimed at the SCN, 0.3 mm anterior to bregma, 0.20 mm lateral to the midline, and 5.8 mm ventral to the surface of the cortex, according to the Paxinos Mouse Brain Atlas, 2001. Bilateral SCN lesions were made by passing a 1-mA current through the electrode for 6 s, in the left and right SCN separately. Sham-lesioned mice underwent the same procedure, but no current was passed through the electrode. After surgery animals were housed individually under constant dark conditions with ad libitum food and water and were allowed to recover for 1 week before behavioural analysis. Successfully SCN-lesioned mice were selected by magnetic resonance imaging (MRI), arrhythmic behaviour and histopathology analysis.

**Magnetic resonance imaging.** Screening of SCN ablated mice was performed using a Bruker ICON scanner (Bruker, Karlsruhe, Germany). RARE (Rapid Acquisition with Refocused Echoes) sequence was used to acquire coronal, sagittal and axial slices (five slices in each orientation) with the following parameters: RARE factor = 8, TE = 85 ms, TR = 2,500 ms, resolution = $156 \times 156 \times 500\ \mu m^3$ (30 averages). For high-quality images, a 9.4-T BioSpec scanner (Bruker, Karlsruhe, Germany) was used. This operates with Paravision 6.0.1 software and is interfaced with an Avance IIIHD console. Anatomical images (16 axial and 13 sagittal slices) were acquired using a RARE sequence with RARE factor = 8, TE = 36 ms, TR = 2,200 ms and resolution of $80 \times 80 \times 500\ \mu m^3$ (12 averages).

**Behavioural analysis.** Sham-operated and SCN-ablated mice were individually housed and after a 24-h acclimation period their movement was recorded for 72 h, in constant darkness, using the automated animal behaviour CleverSys system. Data were auto scored by the CleverSys software. Videos and scoring were visually validated. Circadian rhythmicity was evaluated using the cosinor regression model[40,41].

**Histopathology analysis.** Mice infected with *C. rodentium* were killed by $CO_2$ narcosis, the gastrointestinal tract was isolated, and the full length of caecum and colon was collected and fixed in 10% neutral buffered formalin. Colon was trimmed in multiple transverse and cross-sections and caecum in one cross-section[42], and all were processed for paraffin embedding. Sections (3–4 μm) were stained with haematoxylin and eosin and lesions were scored by a pathologist blinded to experimental groups, according to previously published criteria[43–45]. In brief, lesions were individually scored (0–4 increasing severity) for: mucosal loss; mucosal epithelial hyperplasia; degree of inflammation; extent of the section affected in any manner; and extent of the section affected in the most severe manner, as previously described[45]. The score was derived by summing the individual lesion and extent scores. Mesenteric (mesocolic) inflammation was noted but not scored. Liver, gonadal and subcutaneous fat from *Arntl^ΔRorgt* mice was collected, fixed in 10% neutral buffered formalin, processed for paraffin embedding, sectioned into 3-μm-thick sections and stained with haematoxylin and eosin. The presence of inflammatory infiltrates was analysed by a pathologist blinded to experimental groups. For the SCN lesions experiment, mice were killed with $CO_2$ narcosis, necropsy was performed and brain was harvested and fixed in 4% PFA. Coronal sections of 50-μm thickness were prepared with a vibratome (Leica VT1000 S), from 0.6 to –1.3 relative to the bregma, collected on Superfrost Plus slides (Menzel-Gläser) and allowed to dry overnight before Nissl staining. Stained slides were hydrated in distilled water for a few seconds and incubated in Cresyl Violet stain solution (Sigma-Aldrich) for 30 min. Slides were dehydrated in graded ethanol and mounted with CV Mount (Leica). Coronal sections were analysed for the presence or absence of an SCN lesion (partial versus total ablation, unilateral versus bilateral) in a Leica DM200 microscope coupled to a Leica MC170HD camera (Leica Microsystems, Wetzlar, Germany).

**Microscopy.** Adult intestines from *Ret^GFP* mice were flushed with cold PBS (Gibco) and opened longitudinally. Mucus and epithelium were removed, and intestines were fixed in 4% PFA (Sigma-Aldrich) at room temperature for 10 min and incubated in blocking/permeabilizing buffer solution (PBS containing 2% BSA, 2% goat serum, 0.6% Triton X-100). Samples were cleared with benzyl alcohol-benzyl benzoate (Sigma-Aldrich) prior to dehydration in methanol[18,46]. Whole-mount samples were incubated overnight or for 2 days at 4 °C using the following antibodies: anti-tyrosine hydroxylase (TH) (Pel-Freez Biologicals) and anti-GFP (Aves Labs). Alexa Fluor 488 goat anti-chicken and Alexa Fluor 568 goat anti-rabbit (Invitrogen) were used as secondary antibodies overnight at room temperature. For SCN imaging, *RFP^ΔCamk2a* and *RFP^ΔRorgt* mice were anaesthetized and perfused intracardially with PBS followed by 4% paraformaldehyde (pH 7.4, Sigma-Aldrich). The brains were removed and post-fixed for 24 h in 4% paraformaldehyde and transferred to phosphate buffer. Coronal sections (50 μm) were collected through the entire SCN using a Leica vibratome (VT1000s) into phosphate buffer and processed free-floating. Sections were incubated with neurotrace 500/525 (Invitrogen, N21480) diluted 1/200 and mounted using Mowiol. Samples were acquired on a Zeiss LSM710 confocal microscope using EC Plan-Neofluar 10×/0.30 M27, Plan Apochromat 20×/0.8 M27 and EC Plan-Neofluar 40×/1.30 objectives.

**RNA sequencing and data analysis.** RNA was extracted and purified from sorted small intestinal lamina propria cells isolated at ZT5 and ZT23. RNA quality was assessed using an Agilent 2100 Bioanalyzer. SMART-SeqII (ultra-low input RNA) libraries were prepared using Nextera XT DNA sample preparation kit (Illumina). Sequencing was performed on an Illumina HiSeq4000 platform, PE100. Global quality of FASTQ files with raw RNA-seq reads was analysed using fastqc (ver 0.11.5) (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Vast-tools[47] (version 2.0.0) aligning and read processing software was used for quantification of gene expression in read counts from FASTQ files using VASTD-DB[47] transcript annotation for mouse genome assembly mm9. Only the 8,443 genes with read count information in all 12 samples and an average greater than 1.25 reads per sample were considered informative enough for subsequent analyses. Preprocessing of read count data, namely transforming them to $\log_2$(counts per million) (logCPM), was performed with voom[48], included in the Bioconductor[49] package limma[50] (version 3.38.3) for the statistical software environment R (version 3.5.1). Linear models and empirical Bayes statistics were used for differential gene expression analysis, using limma. For heat maps, normalized RNA-seq data were plotted using the pheatmap (v1.0.10) R package (http://www.R-project.org/). Heat-map genes were clustered using Euclidean distance as metric.

**Statistics.** Results are shown as mean ± s.e.m. Statistical analysis was performed using GraphPad Prism software (version 6.01). Comparisons between two samples were performed using Mann–Whitney $U$ test or unpaired Student's $t$-test. Two-way ANOVA analysis was used for multiple group comparisons, followed by Tukey's post hoc test or Sidak's multiple comparisons test. Circadian rhythmicity was evaluated using the cosinor regression model[40,41,51], using the cosinor (v1.1) R package. A single-component cosinor fits one cosine curve by least squares to the data. The circadian period was assumed to be 24 h for all analysis and the significance of the circadian fit was assessed by a zero-amplitude test with 95% confidence. A

single-component cosinor yields estimates and defines standard errors with 95% confidence limits for amplitude and acrophase using Taylor's series expansion[51]. The latter were compared using two-tailed Student's $t$-test where indicated. Results were considered significant at $*P < 0.05$, $**P < 0.01$, $***P < 0.001$.

**Reporting summary.** Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

## Data availability

Source data for quantifications shown in all graphs plotted in the Figures and Extended Data Figures are available in the online version of the paper. The datasets generated in this study are also available from the corresponding author upon reasonable request. RNA-seq datasets analysed are publicly available in the Gene Expression Omnibus repository with accession number GSE135235.

21. Mombaerts, P. et al. RAG-1-deficient mice have no mature B and T lymphocytes. *Cell* **68**, 869–877 (1992).
22. Cao, X. et al. Defective lymphoid development in mice lacking expression of the common cytokine receptor gamma chain. *Immunity* **2**, 223–238 (1995).
23. Shinkai, Y. et al. RAG-2-deficient mice lack mature lymphocytes owing to inability to initiate V(D)J rearrangement. *Cell* **68**, 855–867 (1992).
24. de Boer, J. et al. Transgenic mice with hematopoietic and lymphoid specific expression of Cre. *Eur. J. Immunol.* **33**, 314–325 (2003).
25. Sawa, S. et al. Lineage relationship analysis of RORγt+ innate lymphoid cells. *Science* **330**, 665–669 (2010).
26. Casanova, E. et al. A CamKIIα iCre BAC allows brain-specific gene inactivation. *Genesis* **31**, 37–42 (2001).
27. Schlenner, S. M. et al. Fate mapping reveals separate origins of T cells and myeloid lineages in the thymus. *Immunity* **32**, 426–436 (2010).
28. Cheng, H. Y. et al. Segregation of expression of mPeriod gene homologs in neurons and glia: possible divergent roles of mPeriod1 and mPeriod2 in the brain. *Hum. Mol. Genet.* **18**, 3110–3124 (2009).
29. Hoshi, M., Batourina, E., Mendelsohn, C. & Jain, S. Novel mechanisms of early upper and lower urinary tract patterning regulated by RetY1015 docking tyrosine in mice. *Development* **139**, 2405–2415 (2012).
30. Madisen, L. et al. A robust and high-throughput Cre reporting and characterization system for the whole mouse brain. *Nat. Neurosci.* **13**, 133–140 (2010).
31. Chomez, P. et al. Increased cell death and delayed development in the cerebellum of mice lacking the rev-erbA(alpha) orphan receptor. *Development* **127**, 1489–1498 (2000).
32. Storch, K. F. et al. Intrinsic circadian clock of the mammalian retina: importance for retinal processing of visual information. *Cell* **130**, 730–741 (2007).
33. Mittelstadt, P. R., Monteiro, J. P. & Ashwell, J. D. Thymocyte responsiveness to endogenous glucocorticoids is required for immunological fitness. *J. Clin. Invest.* **122**, 2384–2394 (2012).
34. Hinoi, E. et al. The sympathetic tone mediates leptin's inhibition of insulin secretion by modulating osteocalcin bioactivity. *J. Cell Biol.* **183**, 1235–1242 (2008).
35. Fonseca-Pereira, D. et al. The neurotrophic factor receptor RET drives haematopoietic stem cell survival and function. *Nature* **514**, 98–101 (2014).
36. Wiles, S., Pickard, K. M., Peng, K., MacDonald, T. T. & Frankel, G. In vivo bioluminescence imaging of the murine pathogen *Citrobacter rodentium*. *Infect. Immun.* **74**, 5391–5396 (2006).
37. Collins, J. W. et al. *Citrobacter rodentium*: infection, inflammation and the microbiota. *Nat. Rev. Microbiol.* **12**, 612–623 (2014).
38. Rakoff-Nahoum, S., Paglino, J., Eslami-Varzaneh, F., Edberg, S. & Medzhitov, R. Recognition of commensal microflora by toll-like receptors is required for intestinal homeostasis. *Cell* **118**, 229–241 (2004).
39. Druzd, D. et al. Lymphocyte circadian clocks control lymph node trafficking and adaptive immune responses. *Immunity* **46**, 120–132 (2017).
40. Henslee, E. A. et al. Rhythmic potassium transport regulates the circadian clock in human red blood cells. *Nat. Commun.* **8**, 1978 (2017).
41. Refinetti, R., Lissen, G. C. & Halberg, F. Procedures for numerical analysis of circadian rhythms. *Biol. Rhythm Res.* **38**, 275–325 (2007).
42. Moolenbeek, C. & Ruitenberg, E. J. The "Swiss roll": a simple technique for histological studies of the rodent intestine. *Lab. Anim.* **15**, 57–59 (1981).
43. Burich, A. et al. Helicobacter-induced inflammatory bowel disease in IL-10- and T cell-deficient mice. *Am. J. Physiol. Gastrointest. Liver Physiol.* **281**, G764–G778 (2001).
44. Fort, M. M. et al. A synthetic TLR4 antagonist has anti-inflammatory effects in two murine models of inflammatory bowel disease. *J. Immunol.* **174**, 6416–6423 (2005).
45. Seamons, A., Treuting, P. M., Brabb, T. & Maggio-Price, L. Characterization of dextran sodium sulfate-induced inflammation and colonic tumorigenesis in *Smad3*−/− mice with dysregulated TGFβ. *PLoS One* **8**, e79182 (2013).
46. Veiga-Fernandes, H. et al. Tyrosine kinase receptor RET is a key regulator of Peyer's patch organogenesis. *Nature* **446**, 547–551 (2007).
47. Tapial, J. et al. An atlas of alternative splicing profiles and functional associations reveals new regulatory programs and genes that simultaneously express multiple major isoforms. *Genome Res.* **27**, 1759–1768 (2017).
48. Law, C. W., Chen, Y., Shi, W. & Smyth, G. K. voom: Precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biol.* **15**, R29 (2014).
49. Huber, W. et al. Orchestrating high-throughput genomic analysis with Bioconductor. *Nat. Methods* **12**, 115–121 (2015).
50. Ritchie, M. E. et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **43**, e47 (2015).
51. Cornelissen, G. Cosinor-based rhythmometry. *Theor. Biol. Med. Model.* **11**, 16 (2014).

**Author contributions** C.G.-S. and R.G.D. designed, performed and analysed the experiments shown in Figs. 1–4 and Extended Data Figs. 1–10. M.R. performed circadian analysis. M.R. and H.R. provided technical assistance for the experiments shown in Fig. 2d–j and Extended Data Fig. 5f–h. M.R. provided technical assistance for the experiments shown in Fig. 3b and Extended Data Fig. 6b, e. B.R. provided technical assistance for the experiments shown in Figs. 1a–d, 3a, h, i, 4j–k and Extended Data Figs. 1, 3f, j, 6a, k, l, 8d, 9e–h. H.R. managed the animal colony. J.A.d.S. and R.M.C. helped to design the experiments shown in Fig. 4e and Extended Data Fig. 8b. A.V. provided technical assistance with flow cytometry. N.L.B.-M. analysed the experiments shown in Fig. 3b and Extended Data Fig. 6b. T.C. analysed the experiments in Fig. 2e, f and Extended Data Figs. 5a–c, 8d. H.V.-F. supervised the work, planned the experiments and wrote the manuscript.
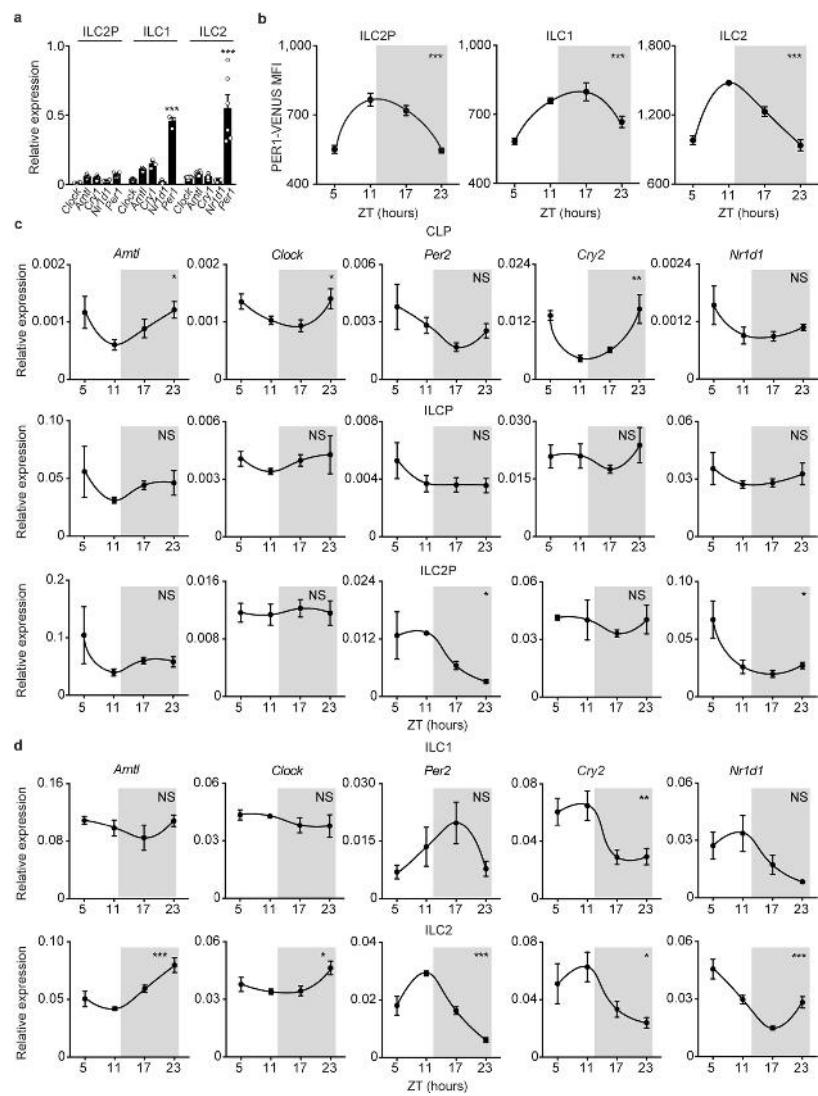
**Extended Data Fig. 1 | Clock genes in progenitors and gut ILCs.**
**a**, Circadian clock gene expression in bone marrow ILC2Ps and small intestinal ILC1s and ILC2s. ILC2P $n = 4$; ILC1 $n = 3$; ILC2 $n = 6$.
**b**, PER1–VENUS MFI analysis in bone marrow ILC2Ps and small intestine lamina propria ILC1s and ILC2s. ILC2P $n = 6$; ILC1 and ILC2 $n = 4$.
**c**, Circadian clock gene expression in bone marrow CLPs, ILCPs and ILC2Ps; $n = 3$. **d**, Circadian clock gene expression in small intestine lamina propria ILC1s and ILC2s. ILC1 $n = 3$; ILC2 $n = 6$. **b–d**, White, light period; grey, dark period. Data are representative of three independent experiments. **a**, **c**, **d**, $n$ represents biologically independent samples; **b**, $n$ represents biologically independent animals. Data shown as mean $\pm$ s.e.m. **a**, Two-way ANOVA followed by Tukey's multiple comparison test. $P$ values relative to differences in *Per1* expression in ILC1s and ILC2s when compared with ILC2Ps. **b–d**, Cosinor regression was used to define circadian rhythmicity. $*P < 0.05$; $**P < 0.01$; $***P < 0.001$; NS, not significant.
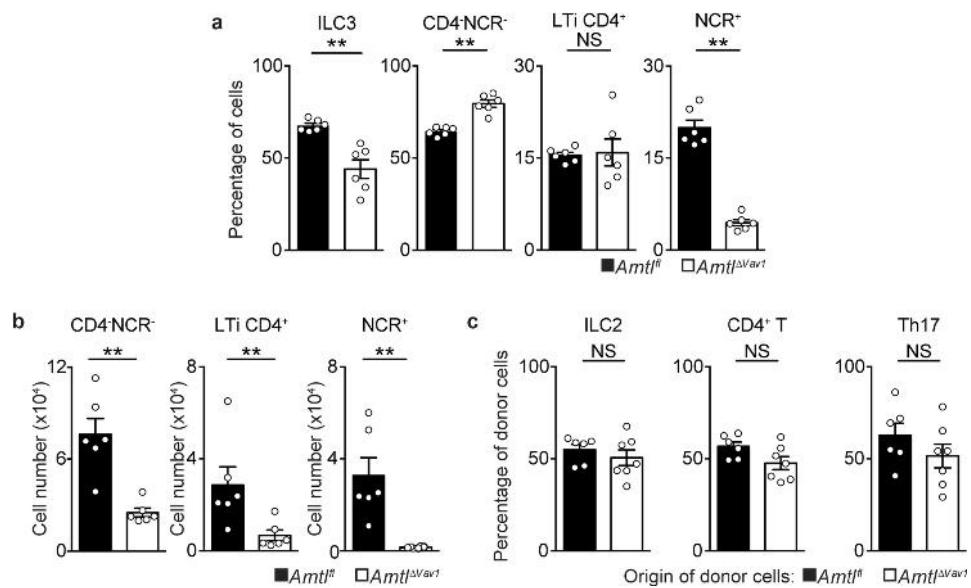
**Extended Data Fig. 2 | Circadian signals regulate gut ILC3. a,** Percentage of lamina propria ILC3s; CD4−NCR−, LTi CD4+ and NCR+ subsets; $n = 6$. **b,** Cell numbers of lamina propria group 3 ILCs; CD4−NCR−, LTi CD4+ and NCR+ ILC3 subsets; $n = 6$. **c,** Percentage of donor cells in ILC2s, CD4+ T cells and Th17 subsets of mixed bone marrow chimaeras. $Arntl^{fl}$ $n = 6$, $Arntl^{\Delta Vav1}$ $n = 7$. Data are representative of three independent experiments. **a–c,** $n$ represents biologically independent animals. Data shown as mean ± s.e.m. **a–c,** Two-tailed Mann–Whitney $U$ test. **$P < 0.01$; NS, not significant.
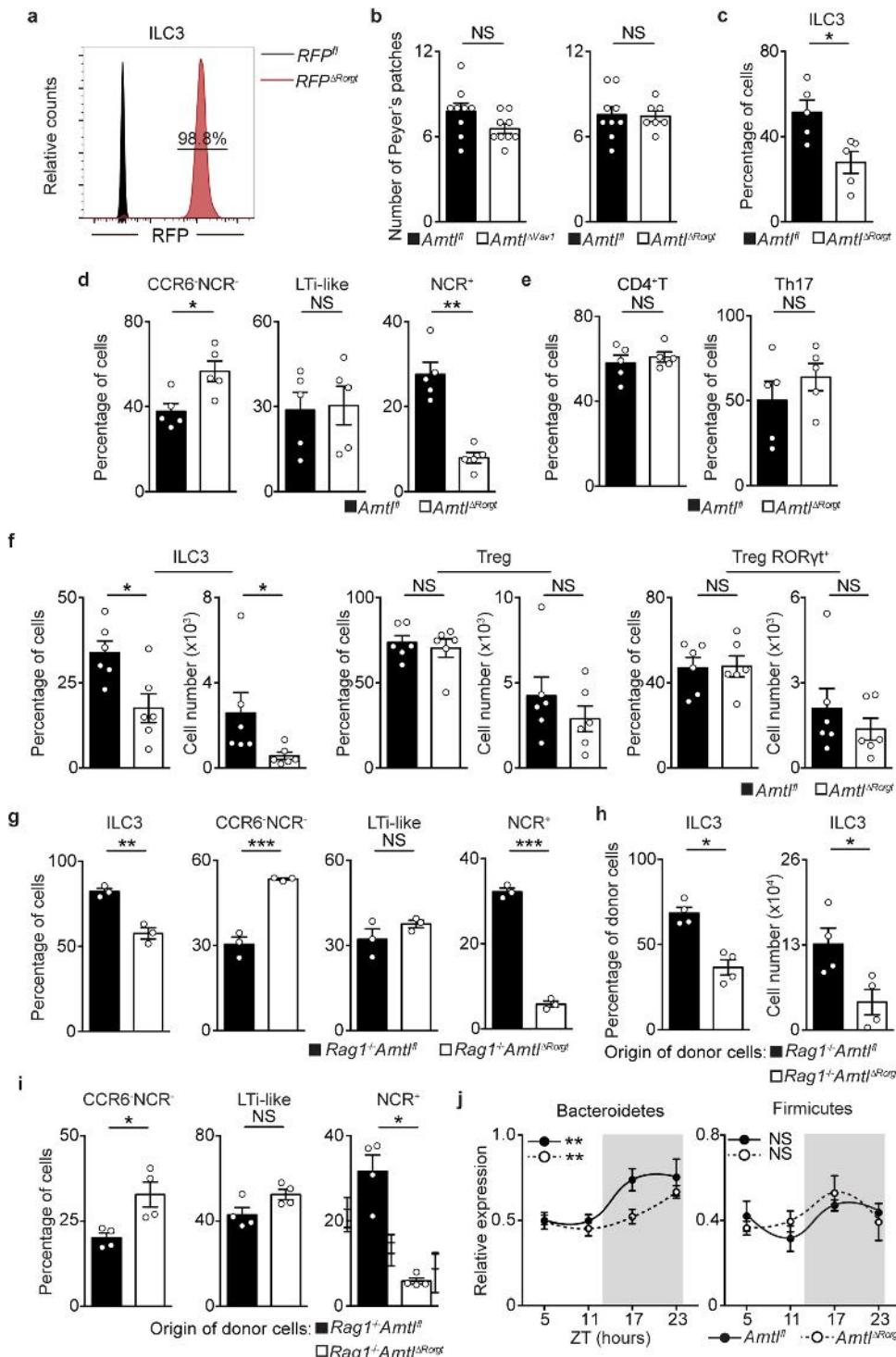
**Extended Data Fig. 3 | Cell-intrinsic *Arntl* signals control intestinal ILC3s. a**, Representative histogram of RFP expression in small intestine lamina propria ILC3s. Representative of three independent analyses. **b**, Number of Peyer patches. *Arntl^fl* and *Arntl^{ΔVav1}* $n = 9$; *Arntl^fl* and *Arntl^{ΔRorgt}* $n = 7$. **c**, Percentage of small intestine lamina propria ILC3s; $n = 5$. **d**, Percentage of CCR6⁻NCR⁻, CCR6⁺ (LTi-like) and NCR⁺ ILC3 subsets; $n = 5$. **e**, Percentage of intestinal lamina propria CD4⁺ T cells and T helper (Th)17 cells; $n = 5$. **f**, Percentage and cell numbers of ILC3s, regulatory T cells (Tregs) and ROR-γt-expressing Tregs in the colonic lamina propria; $n = 6$. **g**, Percentage of lamina propria ILC3s, CCR6⁻NCR⁻, CCR6⁺ (LTi-like) and NCR⁺ ILC3 subsets; $n = 3$.

**h**, Percentage of donor cells and cell numbers of intestinal ILC3s in mixed bone marrow chimaeras; $n = 4$. **i**, Percentage of donor cells in CCR6⁻NCR⁻, CCR6⁺ (LTi-like), and NCR⁺ ILC3 subsets of mixed bone marrow chimaeras; $n = 4$. **j**, Faecal Bacteroidetes and Firmicutes. *Arntl^fl* $n = 5$; *Arntl^{ΔRorgt}* $n = 6$. **j**, White, light period; grey, dark period. Data are representative of at least three independent experiments. **a–j**, $n$ represents biologically independent animals. Data shown as mean ± s.e.m. **b–f, h–i**, Two-tailed Mann–Whitney $U$ test; **j**, cosinor regression was used to define circadian rhythmicity; **g**, two-tailed unpaired Student's $t$-test. *$P < 0.05$; **$P < 0.01$; ***$P < 0.001$; NS, not significant.
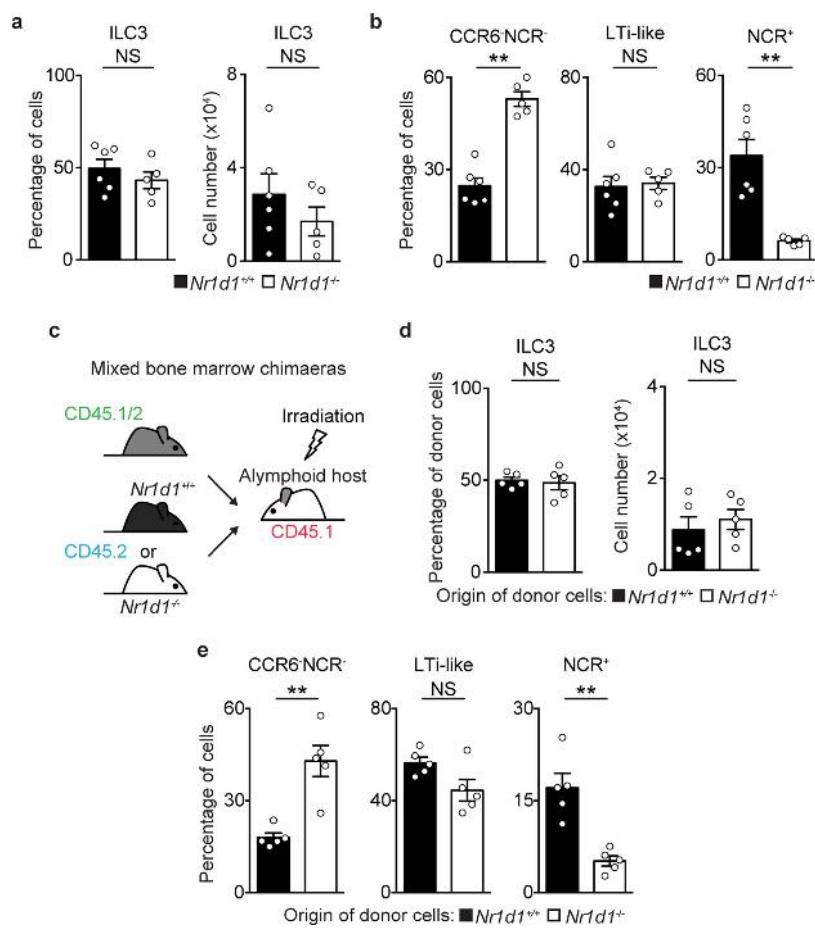
**Extended Data Fig. 4 | Effect of *Nr1d1* in intestinal ILC3s. a**, Percentage and cell numbers of small intestine lamina propria ILC3s. $N1d1^{+/+}$ $n = 6$; $N1d1^{-/-}$ $n = 5$. **b**, Percentage of CCR6−NCR−, CCR6+ (LTi-like) and NCR+ ILC3 subsets. $N1d1^{+/+}$ $n = 6$; $N1d1^{-/-}$ $n = 5$. **c**, Schematic representation of mixed bone marrow chimaeras. **d**, Percentage and cell numbers of donor cells in mixed bone marrow chimaeras; $n = 5$.

**e**, Percentage of CCR6−NCR−, CCR6+ (LTi-like), and NCR+ ILC3 subsets in the small intestine lamina propria of mixed bone marrow chimaeras; $n = 5$. Data are representative of at least three independent experiments. *n* represents biologically independent animals. Data shown as mean ± s.e.m. **a**, **b**, **d**, **e**, Two-tailed Mann–Whitney *U* test. \*\*$P < 0.01$; NS, not significant.
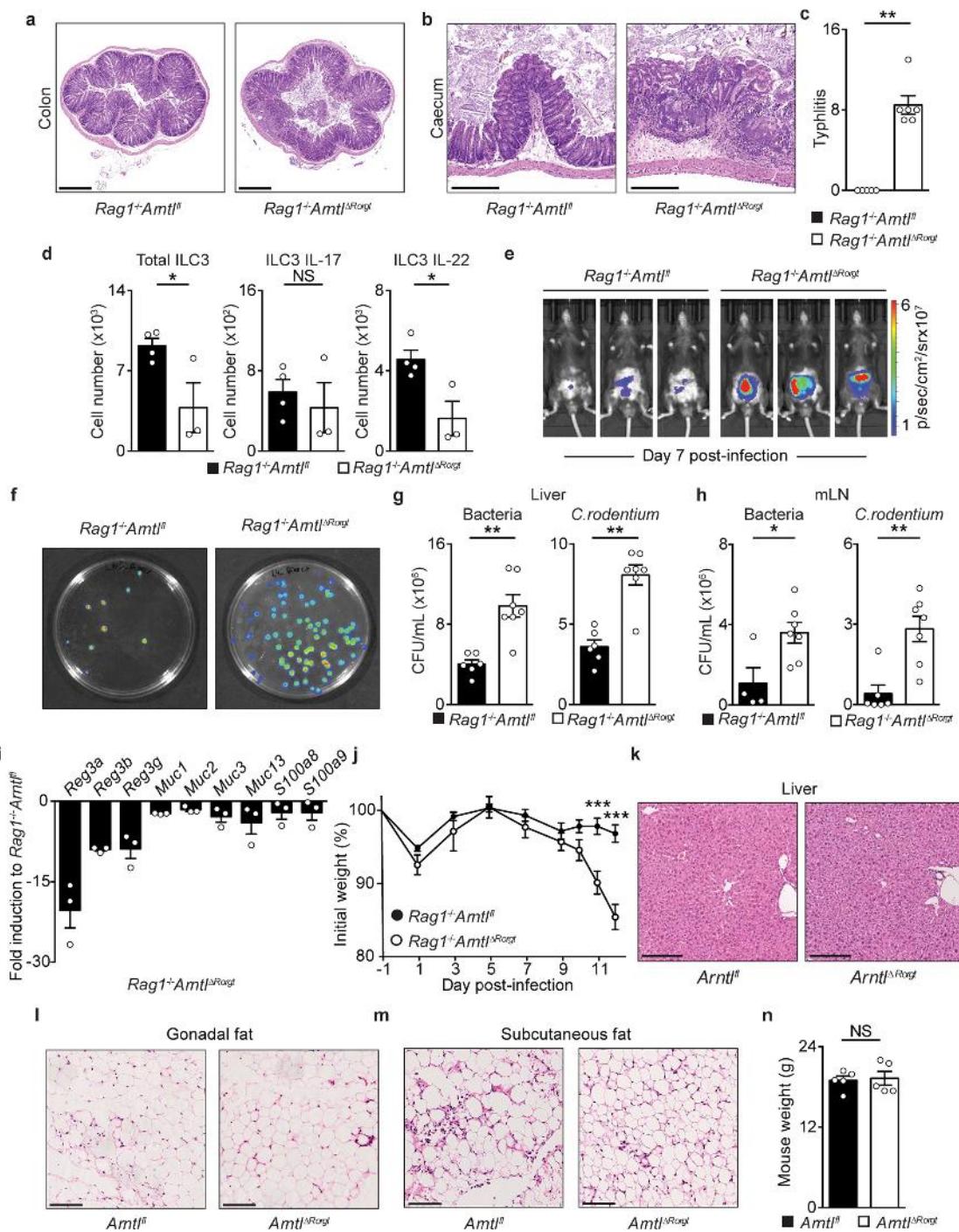
**Extended Data Fig. 5 | ILC3-autonomous ablation of *Arntl* impairs intestinal defence. a**, **b**, Histopathology of colon and caecum in *Rag1*[−/−]*Arntl*[ΔRorgt] and *Rag1*[−/−]*Arntl*[fl] littermate controls infected with *C. rodentium*. Pathological changes in colon and caecum of *Rag1*[−/−]*Arntl*[ΔRorgt] mice included ulceration, loss of crypts and goblet cells, and inflammatory cell infiltration of the lamina propria by a granulocyte-rich population with a prominent and oedematous submucosa. Original magnification ×4 (colon: scale bars, 500 μm); ×1.25 (caecum: scale bars, 250 μm). **c**, Inflammation score in the caecum; $n = 5$. **d**, Total number of ILC3s and numbers of IL-17- and IL-22-producing ILC3s; $n = 3$. **e**, Whole-body imaging of *Rag1*[−/−]*Arntl*[ΔRorgt] and their *Rag1*[−/−]*Arntl*[fl] littermate controls at day 7 after infection with luciferase-expressing *C. rodentium*. **f**, MacConkey plates of liver cell suspensions from *Rag1*[−/−]*Arntl*[ΔRorgt] mice and their *Rag1*[−/−]*Arntl*[fl] littermate

controls at day 13 after *C. rodentium* infection. **g**, **h**, Translocation of total bacteria (left) and *C. rodentium* (right) to the liver and mesenteric lymph nodes (mLN); $n = 4$. **i**, Epithelial reactivity gene expression in the colon of *Rag1*[−/−]*Arntl*[ΔRorgt] mice compared with *Rag1*[−/−]*Arntl*[fl] littermate controls infected with *C. rodentium*; $n = 3$. **j**, Weight loss in *C. rodentium*-infected mice. *Rag1*[−/−]*Arntl*[fl] $n = 6$; *Rag1*[−/−]*Arntl*[ΔRorgt] $n = 7$. **k**–**m**, Histopathology analysis of inflammatory infiltrates in the liver and gonadal and subcutaneous fat. Scale bars: 250 μm (liver); 100 μm (gonadal and subcutaneous fat); $n = 4$. **n**, Total body weight; $n = 5$. Data are representative of at least three independent experiments. *n* represents biologically independent animals. Data shown as mean ± s.e.m. **c**, **g**, **h**, Two-tailed Mann–Whitney *U* test; **d**, **n**, two-tailed unpaired Student's *t*-test; **j**, two-way ANOVA and Sidak's test. *$P < 0.05$; **$P < 0.01$; ***$P < 0.001$; NS, not significant.
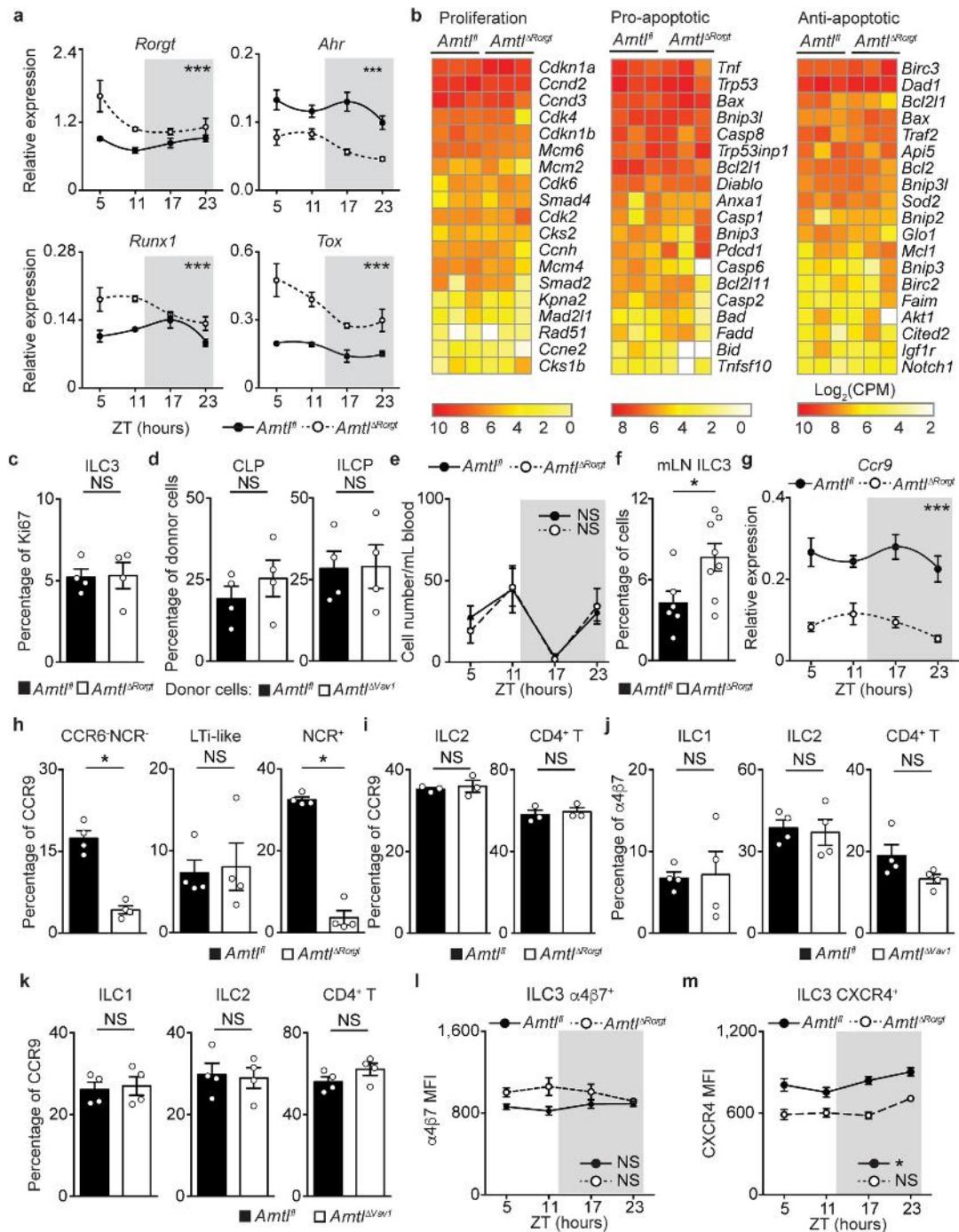
**Extended Data Fig. 6 | ILC3 proliferation, apoptosis and gut homing markers. a**, ILC3-related gene rhythmicity in small intestinal lamina propria ILC3s; $n = 4$. **b**, RNA-seq analysis of lamina propria ILC3s at ZT23; $n = 3$. **c**, Percentage of Ki67 expression in small intestine lamina propria ILC3s; $n = 4$. **d**, Percentage of donor cells in mixed bone marrow chimaeras; $n = 4$. **e**, Number of Lin⁻CD127⁺RORγt⁺ cells in the blood; $n = 4$. **f**, Percentage of ILC3s in mLNs. $Arntl^{fl}$ $n = 6$; $Arntl^{\Delta Rorgt}$ $n = 8$. **g**, Diurnal expression of *Ccr9* transcripts in gut ILC3s; $n = 4$. **h**, **i**, Percentage of CCR9 expression in small intestinal lamina propria CCR6⁻NCR⁻, CCR6⁺ (LTi-like) and NCR⁺ ILC3 subsets, ILC2s and

CD4⁺ T cells; $n = 3$. **j**, Percentage of α4β7 expression in small intestine ILC1s, ILC2s and CD4⁺ T cells; $n = 4$. **k**, Percentage of CCR9 expression in gut ILC1s, ILC2s and CD4⁺ T cells; $n = 4$. **l**, **m**, Diurnal analysis of α4β7 and CXCR4 expression in small intestine ILC3s; $n = 4$. **a**, **e**, **g**, **l**, **m**, White, light period; grey, dark period. Data shown as mean ± s.e.m. **a**, **c**–**m**, $n$ represents biologically independent animals. **b**, $n$ represents biologically independent samples. **a**, **g**, Two-way ANOVA; **c**, **d**, **h**–**k**, two-tailed Mann–Whitney $U$ test; **e**, **l**, **m**, cosinor regression was used to define circadian rhythmicity; **f**, two-tailed unpaired Student's $t$-test. *$P < 0.05$; **$P < 0.01$; ***$P < 0.001$; NS, not significant.
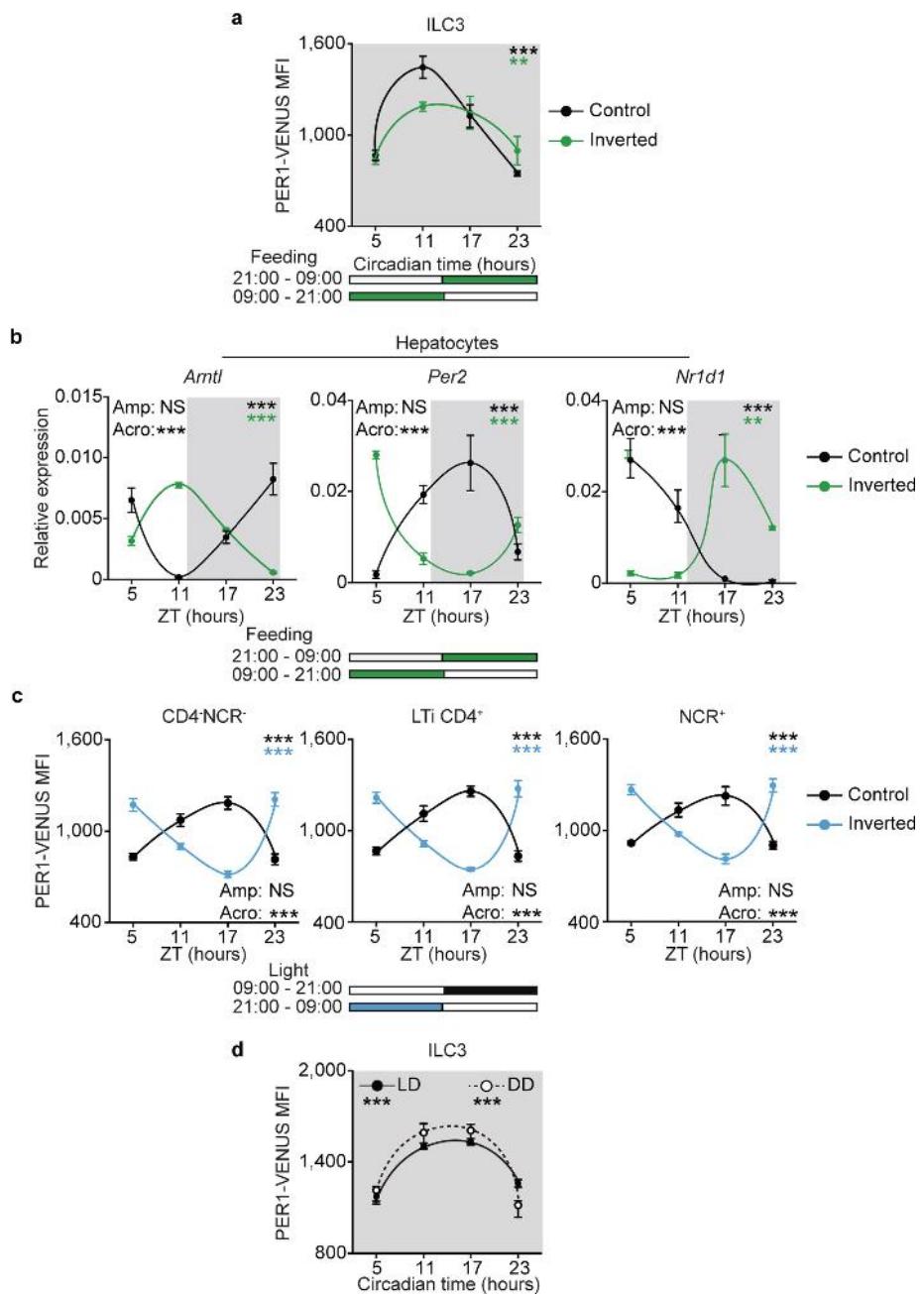
**Extended Data Fig. 7 | Light entrains intestinal ILC3 circadian oscillations. a**, Inverted feeding regimens in constant darkness. PER1–VENUS expression in gut ILC3s; $n = 3$. **b**, Circadian clock gene expression in hepatocytes of $Per1^{Venus}$ mice in inverted feeding regimens; $n = 3$. Acrophase mean $\pm$ s.e.m: $Arntl$: control $0.4 \pm 0.5$, inverted $11.5 \pm 0.2$; $Per2$: control $15.2 \pm 0.6$, inverted $3.9 \pm 0.5$; $Nr1d1$: control $7.1 \pm 0.6$; inverted $18.5 \pm 0.8$. **c**, Opposing light–dark cycles. PER1–VENUS in gut CD4⁻NCR⁻, LTi CD4⁺ and NCR⁺ ILC3 subsets; $n = 3$. Acrophase mean and s.e.m: CD4⁻NCR⁻: control $14.5 \pm 0.5$, inverted $2.5 \pm 0.5$; LTi CD4⁺:

control $14.5 \pm 0.6$, inverted $2.5 \pm 0.4$; NCR⁺: control $14.5 \pm 0.6$, inverted $2.5 \pm 0.4$. **d**, PER1–VENUS MFI analysis of small intestine lamina propria ILC3s from mice maintained in constant darkness for 28 days; $n = 3$. **b**, White, light period; grey, dark period. Data are representative of three independent experiments. $n$ represents biologically independent animals. Data shown as mean $\pm$ s.e.m. Cosinor regression. Standard errors with 95% confidence limits for amplitude (Amp) and acrophase (Acro) were extracted from the model and compared using two-tailed Student's $t$-test. $**P < 0.01$; $***P < 0.001$; NS, not significant.
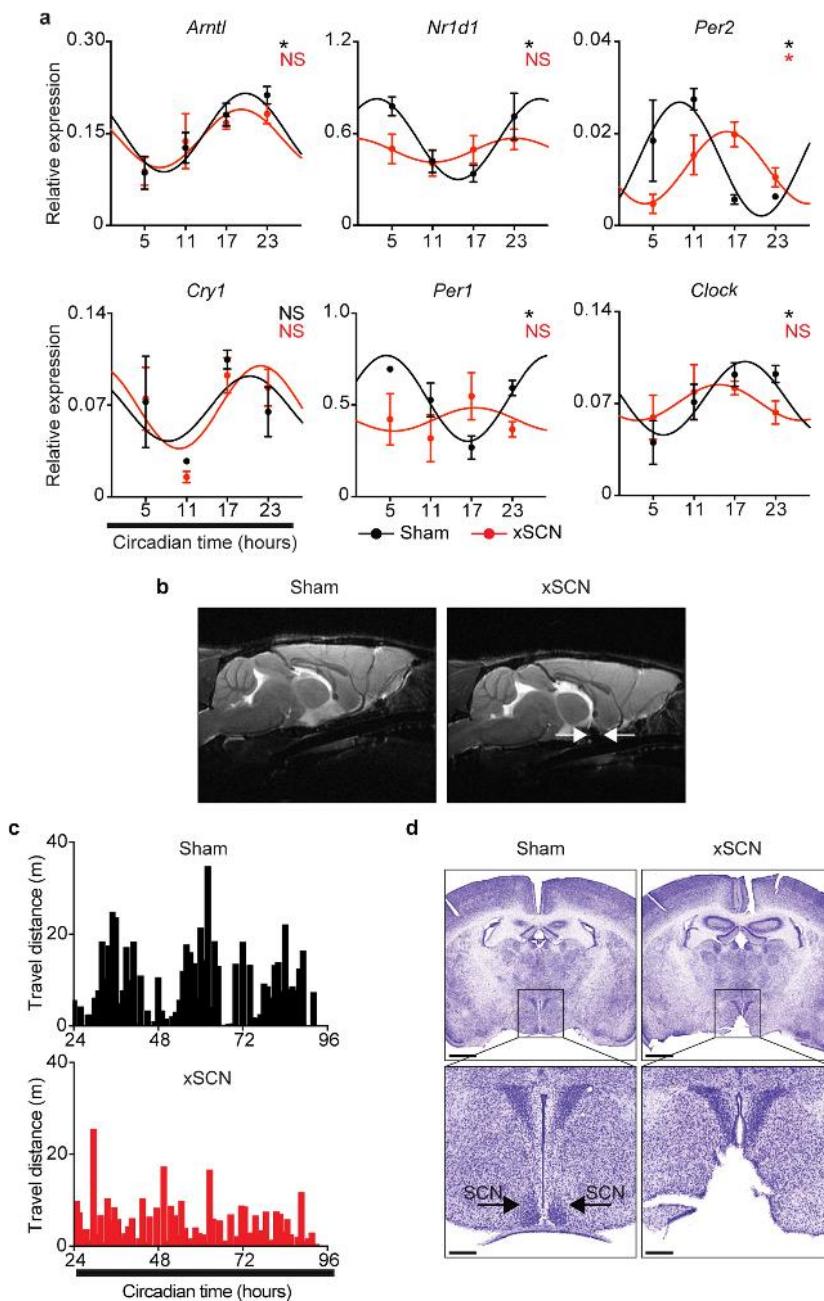
**Extended Data Fig. 8 | SCN ablation shapes intestinal ILC3s.**
**a**, Circadian clock gene expression in small intestinal lamina propria
ILC3s. $n = 2$ or 3. **b**, Magnetic resonance imaging of sham and SCN-
ablated $Per1^{Venus}$ mice. Sagittal slices. White arrows show SCN ablation.
**c**, Rhythms of animal locomotor activity. Total distance travelled in
metres. **d**, Nissl staining of coronal brain sections. Scale bars: 1 mm (top);

250 μm (bottom). $n$ represents biologically independent animals. Data
shown as mean ± s.e.m. Cosinor regression was used to define circadian
rhythmicity; cosine fitted curves are shown; standard errors with 95%
confidence limits for amplitude and acrophase were extracted from the
model and compared using two-tailed Student's $t$-test. *$P < 0.05$; NS, not
significant.
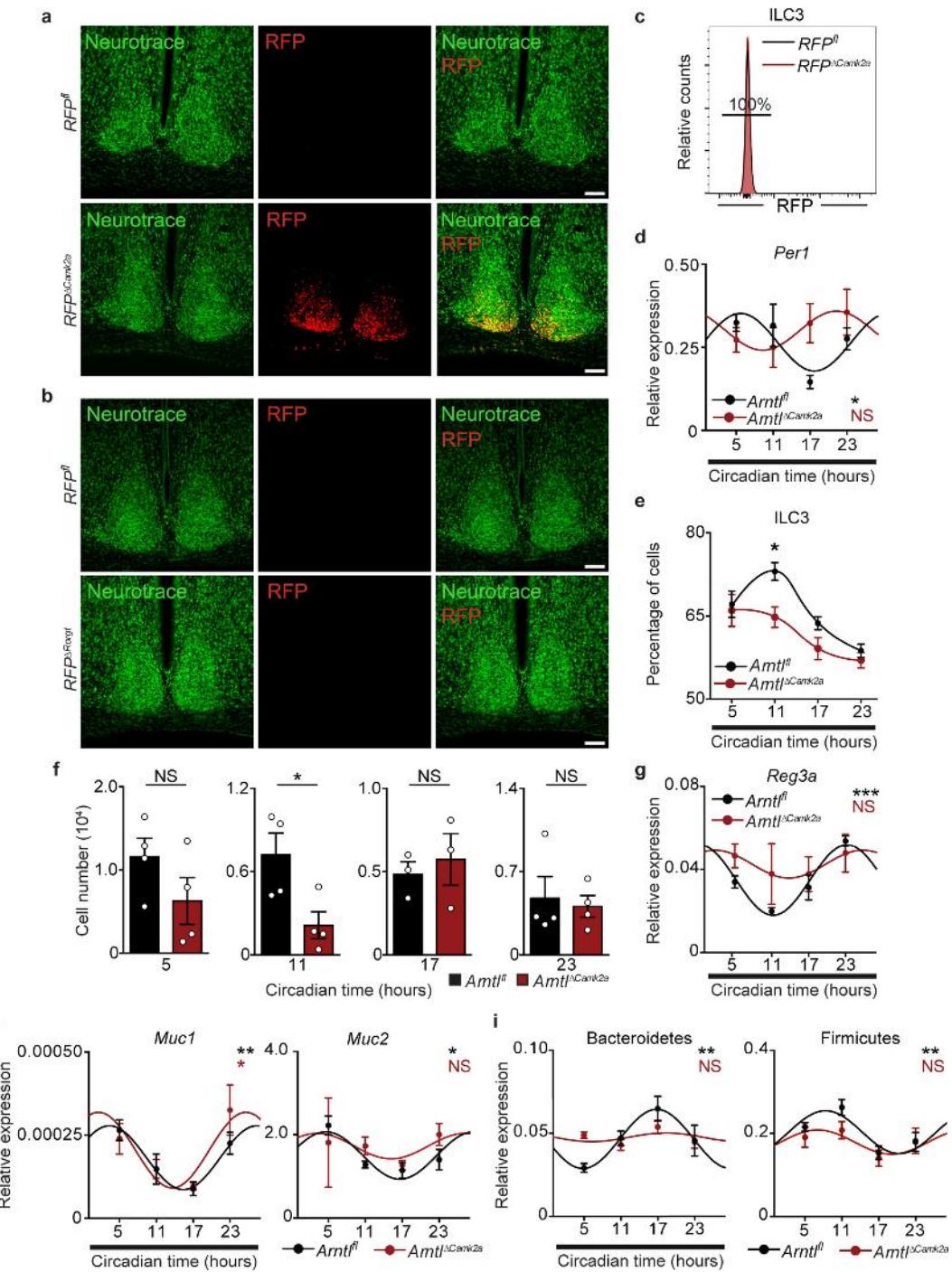
**Extended Data Fig. 9 | Brain-tuned signals shape gut ILC3s.**
**a**, **b**, Confocal images of coronal brain sections showing neurotrace and RFP expression in the SCN. Scale bar, 100 μm. Representative of three independent analyses. **c**, Representative histogram of RFP expression in small intestine lamina propria ILC3s. Representative of three independent analyses. **d**, *Per1* expression in small intestinal lamina propria ILC3s. $n = 3$. **e**, Percentage of small intestine lamina propria ILC3s; $n = 3$. **f**, Number of small intestine lamina propria ILC3s; $n = 3$. **g**, **h**, Epithelial reactivity gene expression in the intestinal epithelium; $n = 3$. **i**, Rhythms of faecal Bacteroidetes and Firmicutes. *Arntl*$^{fl}$ $n = 4$, *Arntl*$^{\Delta Camk2a}$ $n = 3$. $n$ represents biologically independent animals. Data shown as mean ± s.e.m. **d**, **g–i**, Cosinor regression was used to define circadian rhythmicity; cosine fitted curves are shown; **e**, two-way ANOVA and Sidak's test; **f**, two-tailed unpaired Student's *t*-test. *$P < 0.05$; **$P < 0.01$; ***$P < 0.001$; NS, not significant.
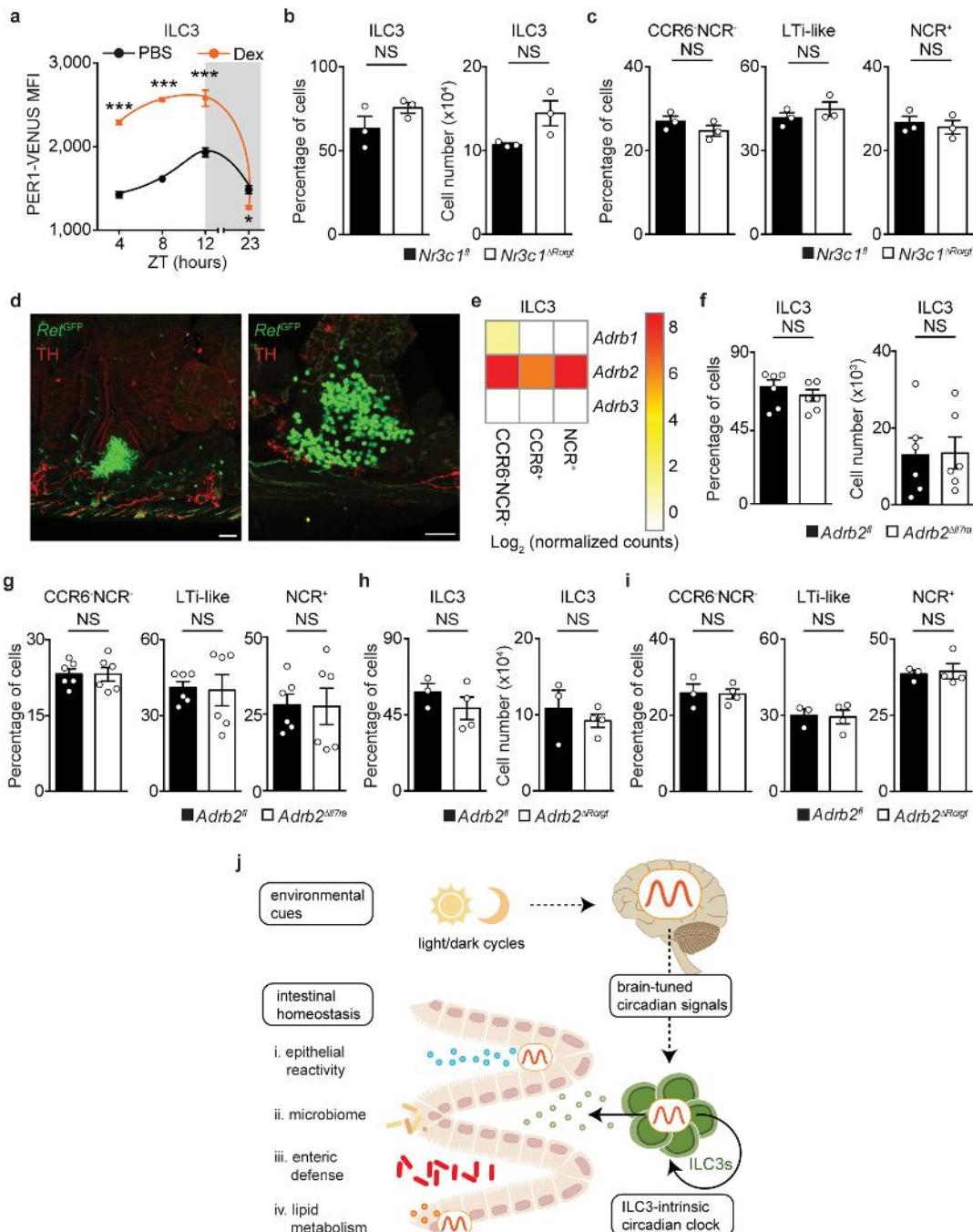
**Extended Data Fig. 10 | Effect of *Nr3c1* and *Adrb2* deficiency in gut ILC3s. a**, PER1–VENUS MFI analysis of lamina propria ILC3s after dexamethasone administration; $n = 3$. **b**, Percentage and cell numbers of small intestine ILC3s; $n = 3$. **c**, Percentage of lamina propria CCR6⁻NCR⁻, CCR6⁺ (LTi-like), and NCR⁺ ILC3 subsets; $n = 3$. **d**, TH-expressing neurons (red) and RET-positive ILC3s (green) in cryptopatches. Scale bars, 40 μm. Representative of three independent analyses. **e**, Normalized expression of *Adrb1*, *Adrb2* and *Adrb3* in CCR6⁻NCR⁻, CCR6⁺ and NCR⁺ ILC3 subsets. **f**, Percentage and cell numbers of gut ILC3s in *Adrb2^{ΔIl7ra}* mice and their littermate controls; $n = 6$. **g**, Percentage of lamina propria CCR6⁻NCR⁻, CCR6⁺ (LTi-like) and NCR⁺ ILC3 subsets in *Adrb2^{ΔIl7ra}* mice and their littermate controls; $n = 6$. **h**, Percentage and cell numbers of gut ILC3s in *Adrb2^{ΔRorgt}* mice and their littermate controls. *Adrb2^{fl}* $n = 3$;

*Adrb2^{ΔRorgt}* $n = 4$. **i**, Percentage of lamina propria CCR6⁻NCR⁻, CCR6⁺ (LTi-like) and NCR⁺ ILC3 subsets in *Adrb2^{ΔRorgt}* mice and their littermate controls. *Adrb2^{fl}* $n = 3$; *Adrb2^{ΔRorgt}* $n = 4$. **j**, Light cues and brain-tuned circuits shape gut ILC3 homeostasis. Arrhythmic ILC3s impact intestinal homeostasis, epithelial reactivity, microbiota, enteric defence and the host lipid metabolism. Thus, ILC3s integrate local and systemic entraining cues in a distinct hierarchical manner, establishing an organismal circuitry that is an essential link between diurnal light signals, brain cues, intestinal ILC3s and host homeostasis. **a**–**d**, **f**–**i**, $n$ represents biologically independent animals. **a**, White, light period; grey, dark period. Data shown as mean ± s.e.m. **a**, Two-way ANOVA and Sidak's test; **b**, **c**, **f**–**i**, two-tailed Mann–Whitney $U$ test. $*P < 0.05$; $***P < 0.001$; NS, not significant.

Corresponding author(s): Veiga-Fernandes

Last updated by author(s): Aug 2, 2019

# Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

## Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

| n/a | Confirmed | |
|---|---|---|
| ☐ | ☒ | The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement |
| ☒ | ☐ | A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly |
| ☐ | ☒ | The statistical test(s) used AND whether they are one- or two-sided <br> *Only common tests should be described solely by name; describe more complex techniques in the Methods section.* |
| ☒ | ☐ | A description of all covariates tested |
| ☐ | ☒ | A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons |
| ☐ | ☒ | A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals) |
| ☐ | ☒ | For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted <br> *Give P values as exact values whenever suitable.* |
| ☒ | ☐ | For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings |
| ☒ | ☐ | For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes |
| ☒ | ☐ | Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated |

*Our web collection on statistics for biologists contains articles on many of the points above.*

## Software and code

Policy information about availability of computer code

| Data collection | FACSAria I, FACSAria III, FACSFusion, LSRFortessa and LSRFortessa X-20 (BD Biosciences); StepOne and QuantStudio 5 Real-Time PCR systems (Applied Biosystems); IVIS Lumina III System (Perkin Elmer); Bruker ICON scanner (Bruker, Karlsruhe, Germany); 9.4T Bruker BioSpec scanner (Bruker, Karlsruhe, Germany) interfaced with an Avance IIIHD console; CleverSys system; Leica DM200 microscope couple to a Leica MC170HD camera (Leica Microsystems, Wetzlar, Germany); Illumina HiSeq4000 platform, PE100. |
|---|---|
| Data analysis | FlowJo 8.8.7 software (Tree Star); StepOne Software v2.3 (Applied Biosystems); QuantStudio Design and Analysis Software v1.4.2 (Applied Biosystems); Living Image v4.5.2 (Caliper Life Sciences); Paravision 6.0.1 software; CleverSys software; pheatmap (v1.0.10) R package; GraphPad Prism software (version 6.01); cosinor (v1.1) R package; fastqc v0.11.5; limma v3.38.3; vast-tools v2.0.0; environment R v3.5.1. |

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

## Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:
- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

The sequencing data generated in this study have been deposited in the GEO repository with the accession number GSE135235.

# Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences     ☐ Behavioural & social sciences     ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

# Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

| | |
|---|---|
| Sample size | Power analysis was performed to estimate the number of experimental mice. |
| Data exclusions | No data exclusions unless stated otherwise. |
| Replication | The experimental findings were reliably reproduced. We further provide thoroughly detailed methods describing the critical steps of each experiment. |
| Randomization | No randomization was performed. Animals were compared with co-housed litter-mate controls. |
| Blinding | Blinding was performed for experiments in Fig.2e,f; Fig.3b; Extended data Fig5a-c,k-m; Extended data Fig6b; Extended data Fig8d. |

# Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

### Materials & experimental systems

| n/a | Involved in the study |
|---|---|
| ☐ | ☒ Antibodies |
| ☒ | ☐ Eukaryotic cell lines |
| ☒ | ☐ Palaeontology |
| ☐ | ☒ Animals and other organisms |
| ☒ | ☐ Human research participants |
| ☒ | ☐ Clinical data |

### Methods

| n/a | Involved in the study |
|---|---|
| ☒ | ☐ ChIP-seq |
| ☐ | ☒ Flow cytometry |
| ☐ | ☒ MRI-based neuroimaging |

## Antibodies

| | |
|---|---|
| Antibodies used | Anti-mouse antibodies used for flow cytometry:<br>anti-CD45 (30-F11; 1:200); anti-CD45.1 (A20; 1:200); anti-CD45.2 (104; 1:200); anti-CD11c (N418; 1:200); anti-CD11b (Mi/70; 1:400); anti-CD127 (IL7Rα; A7R34; 1:100); anti-CD27(LG.7F9; 1:100); anti-CD8α (53-6.7; 1:200); anti-CD19 (eBio1D3; 1:200); anti-CXCR4(L276F12; 1:100); anti-NK1.1 (PK136; 1:100); anti-CD3ε (eBio500A2; 1:200); anti-TER119 (TER-119; 1:200); anti-Gr1 (RB6-8C5; 1:400); anti-CD4 (RM4-5; 1:200); anti-CD25 (PC61; 1:200); anti-CD117 (c-Kit; 2B8; 1:100); anti-CD90.2 (Thy1.2; 53-2.1; 1:200); anti-TCRβ (H57-595; 1:200); anti-TCRγδ (GL3; 1:200); anti-B220 (RA3-6B2; 1:200); anti-KLRG1 (2F1/KLRG1; 1:200); anti-Ly-6A/E (Sca1; D7; 1:200); anti-CCR9 (CW-1.2; 1:100); anti-IL-17 (TC11-18H10.1; 1:100); anti-rat IgG1k isotype control (RTK2071; 1:100); anti-streptavidin fluorochrome conjugates from Biolegend; anti-α4β7 (DATK32; 1:50); anti-Flt3 (A2F10; 1:100); anti-NKp46 (29A1.4; 1:100); anti-CD49b (DX5; 1:100); anti-Ki67 (SolA15; 1:200); anti-rat IgG2ak isotype control (eBR2a; 1:200); anti-IL-22 (1H8PWSR; 1:50); anti-rat IgG1k isotype control (eBRG1; 1:50); anti-EOMES (Dan11mag; 1:200); anti-Tbet (eBio4B10; 1:200); anti-FOPX3 (FJK-16s; 1:100); anti-GATA3 (TWAJ; 1:100); anti-CD16/CD32 (93; 1:50); 7AAD viability dye (1:100) from eBiosciences; anti-CD196 (CCR6; 140706; 1:100) from BD Biosciences; anti-RORgt (Q31-378; 1:100) and anti-mouse IgG2ak isotype control (G155-178; 1:100) from BD Pharmingen. LIVE/DEAD Fixable Aqua Dead Cell Stain Kit (1:50) was purchased from Invitrogen.<br>Antibodies used for microscopy:<br>Anti-Tyrosine hydroxylase (TH; 1:500) (Pel-Freez Biologicals) and anti-GFP (1:1000) (Aves Labs). Alexa Fluor 488 goat anti-chicken (1:300) and Alexa Fluor 568 goat anti-rabbit (1:300) from Invitrogen were used as secondary antibodies. |
| Validation | All the antibodies used in this study were optimized and validated (i.e. assay and species) by manufactures and used according to supplied instructions. Most antibodies were re-validated and appropriate dilutions were determined by titration on ex vivo naive splenocytes. |

# Animals and other organisms

| | |
|---|---|
| Laboratory animals | C57BL/6J mice were purchased from Charles River. Nod/Scid/Gamma (NSG) were purchased from the Jackson Laboratories. C57BL/6J Ly5.1 were purchased from the Jackson Laboratories and bred with C57BL/6J in order to obtain C57BL/6 Ly5.1/Ly5.2 (CD45.1/CD45.2). Rag1-/-, Rag2-/-Il2rg-/-, Vav1Cre, RorgtCre , Camk2aCre, Arntlfl, Per1Venus, Il7ra, RetGFP, Rosa26RFP, Nr1d1-/-, Nr3c1fl and Adrb2fl mouse lines were on a full C57BL/6J background. All lines were bred and maintained at Champalimaud Centre for the Unknown (CCU) animal facility under specific pathogen free conditions. 8-14 weeks old males and females were used in this study. Mice were systematically compared with co-housed littermate controls unless stated otherwise. |
| Wild animals | The study did not involve wild animals. |
| Field-collected samples | The study did not involve field-collected samples. |
| Ethics oversight | All animal experiments were approved by national and institutional ethical committees, respectively, Direção Geral de Veterinária and Champalimaud Centre for the Unkown ethical committees. |

Note that full information on the approval of the study protocol must also be provided in the manuscript.

# Flow Cytometry

## Plots

Confirm that:

☒ The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).

☒ The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).

☒ All plots are contour plots with outliers or pseudocolor plots.

☒ A numerical value for number of cells or percentage (with statistics) is provided.

## Methodology

| | |
|---|---|
| Sample preparation | Intestines and colons were thoroughly rinsed with cold PBS1X, Peyer patches were removed, and intestines and colons were cut in 1cm pieces, and shaken for 30min in PBS containing 2% FBS, 1% HEPES and 5mM EDTA to remove intraepithelial and epithelial cells. Intestines and colons were then digested with collagenase D (0.5mg/mL; Roche) and DNase I (20U/mL; ROCHE) in complete RPMI for 30 min at 37 °C, under gentle agitation. Sequentially cells were passed through a 100μM cell strainer and purified by centrifugation 30min at 2400rpm in 40/80 Percoll (GE Healthcare) gradient. Lungs were finely minced and digested in complete RPMI supplemented with collagenase D (0.1mg/mL; Roche) and DNase I (20U/mL; ROCHE) for 1h at 37°C under gentle agitation. Sequentially, cells were passed through a 100μM cell strainer purified by centrifugation 30 minutes at 2400rpm in 40/80 Percoll (GE Healthcare) gradient. Spleen and mesenteric lymph node cell suspensions were obtained using 70μm strainers. Bone marrow cells were collected by either flushing or crushing bones and filtered using 70μm strainers. Erythrocytes from small intestine, colon, lung, spleen and bone marrow preparations were lysed with RBC lysis buffer (eBioscience). Leukocytes from blood were isolated by treatment with Ficoll (GE Healthcare). |
| Instrument | FACSAria I, FACSAria III, FACSFusion, LSRFortessa and LSRFortessa X-20 (BD Biosciences). |
| Software | FlowJo 8.8.7 software (Tree Star). |
| Cell population abundance | Sorted populations were >95% pure. |
| Gating strategy | Cell populations were defined as: Bone marrow (BM) Common Lymphoid Progenitor (CLP) - Lin-CD127+Flt3+Sca1intc-Kitint; BM Innate Lymphoid Cell Progenitor ILCP - Lin-CD127+Flt3-CD25-c-Kit+α4β7high; BM ILC2 progenitor (ILC2P) - Lin-CD127+Flt3-Sca1+CD25+; small intestine NK - CD45+Lin-NK1.1+NKp46+CD27+CD127-EOMES+CD49b+ or CD45+Lin-NK1.1+NKp46+CD27+CD127-CD49b+ ; small intestine ILC1 - CD45+Lin-NK1.1+NKp46+CD27+CD127+CD49b-Tbet+ or CD45+Lin-NK1.1+NKp46+CD27+CD127+CD49b-; small intestine ILC2 - CD45+Lin-Thy1.2+KLRG1+GATA3+ or CD45+Lin-Thy1.2+KLRG1+SCA1+CD25+; small inestine and colonic lamina propria, lung, spleen and mesenteric lymph node ILC3 - CD45+Lin-Thy1.2hiRORγt+ or CD45+Lin-Thy1.2hiKLRG1-; ILC3 IL17+ - CD45+Lin-Thy1.2hiRORγt+IL17+; ILC3 IL22+ - CD45+Lin-Thy1.2hiRORγt+IL22+; for ILC3 subsets additional markers were employed: ILC3 NCR-CD4- - NKp46-CD4-; ILC3 LTi CD4+ - NKp46-CD4+; ILC3 CCR6-NCR- - CCR6-NKp46-; ILC3 LTi-like - CCR6+NKp46-; ILC3 NCR+ - NKp46+; small intestine T helper (Th) 17 cells - CD45+Lin+Thy1.2+CD4+RORγt+.; colon regulatory T cells (Tregs) - CD45+, CD3+, Thy1.2+, CD4+, CD25+, FOXP3+ and colon Tregs RORγt+ - CD45+, CD3+, Thy1.2+, CD4+, CD25+, FOXP3+RORγt+. The lineage cocktail for BM, lung, lamina propria, spleen and mesenteric lymph nodes included CD3ε, CD8α, CD19, B220, CD11c, CD11b, Ter119, Gr1, TCRβ, TCRγδ and NK1.1. For NK and ILC1 staining in the small intestine NK1.1 and CD11b were not added to the lineage cocktail. |

☐ Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.

# Magnetic resonance imaging

## Experimental design

| | |
|---|---|
| Design type | Resting state. |
| Design specifications | Not applicable. |
| Behavioral performance measures | Not applicable. |

## Acquisition

| | |
|---|---|
| Imaging type(s) | Anatomical images. |
| Field strength | 9.4 Tesla. |
| Sequence & imaging parameters | Anatomical images (16 axial and 13 sagittal slices) were acquired using a RARE (Rapid Acquisition with Refocused Echoes) sequence with RARE factor=8, TE=36ms, TR=2200ms and resolution of 80x80x500µm3 (12 averages). |
| Area of acquisition | Whole brain. |

Diffusion MRI ☐ Used ☒ Not used

## Preprocessing

| | |
|---|---|
| Preprocessing software | Not applicable. |
| Normalization | Not applicable. |
| Normalization template | Not applicable. |
| Noise and artifact removal | Not applicable. |
| Volume censoring | Not applicable. |

## Statistical modeling & inference

| | |
|---|---|
| Model type and settings | Not applicable. |
| Effect(s) tested | Not applicable. |

Specify type of analysis: ☒ Whole brain ☐ ROI-based ☐ Both

| | |
|---|---|
| Statistic type for inference (See Eklund et al. 2016) | Not applicable. |
| Correction | Not applicable. |

## Models & analysis

| n/a | Involved in the study |
|---|---|
| ☒ | ☐ Functional and/or effective connectivity |
| ☒ | ☐ Graph analysis |
| ☒ | ☐ Multivariate modeling or predictive analysis |

# LETTER

# FHL1 is a major host factor for chikungunya virus infection

Laurent Meertens[1,13]*, Mohamed Lamine Hafirassou[1], Thérèse Couderc[2], Lucie Bonnet-Madin[1], Vasiliya Kril[1], Beate M. Kümmerer[3], Athena Labeau[1], Alexis Brugier[1], Etienne Simon-Loriere[4], Julien Burlaud-Gaillard[5], Cécile Doyen[6], Laura Pezzi[7], Thibaud Goupil[2], Sophia Rafasse[2], Pierre-Olivier Vidalain[8], Anne Bertrand-Legout[9], Lucie Gueneau[9], Raul Juntas-Morales[10], Rabah Ben Yaou[9], Gisèle Bonne[9], Xavier de Lamballerie[7], Monsef Benkirane[6], Philippe Roingeard[5], Constance Delaugerre[1,11], Marc Lecuit[2,12] & Ali Amara[1,13]*

**Chikungunya virus (CHIKV) is a re-emerging alphavirus that is transmitted to humans by mosquito bites and causes musculoskeletal and joint pain[1,2]. Despite intensive investigations, the human cellular factors that are critical for CHIKV infection remain unknown, hampering the understanding of viral pathogenesis and the development of anti-CHIKV therapies. Here we identified the four-and-a-half LIM domain protein 1 (FHL1)[3] as a host factor that is required for CHIKV permissiveness and pathogenesis in humans and mice. Ablation of FHL1 expression results in the inhibition of infection by several CHIKV strains and o'nyong-nyong virus, but not by other alphaviruses and flaviviruses. Conversely, expression of FHL1 promotes CHIKV infection in cells that do not normally express it. FHL1 interacts directly with the hypervariable domain of the nsP3 protein of CHIKV and is essential for the replication of viral RNA. FHL1 is highly expressed in CHIKV-target cells and is particularly abundant in muscles[3,4]. Dermal fibroblasts and muscle cells derived from patients with Emery–Dreifuss muscular dystrophy that lack functional FHL1[5] are resistant to CHIKV infection. Furthermore, CHIKV infection is undetectable in Fhl1-knockout mice. Overall, this study shows that FHL1 is a key factor expressed by the host that enables CHIKV infection and identifies the interaction between nsP3 and FHL1 as a promising target for the development of anti-CHIKV therapies.**

Several host factors involved in mediating infection with CHIKV have been identified[6–9]; however, none of these factors accounts for the tropism of CHIKV for joint and muscle tissues. To uncover host genes required for CHIKV infection, we performed a genome-wide CRISPR–Cas9 screen in near haploid human HAP1 cells using CHIKV21, a strain isolated from a patient infected during the 2005–2006 CHIKV outbreak in La Reunion Island[10] (Fig. 1a, Extended Data Fig. 1 and Supplementary Table 1). The top hit of our screen was the gene encoding FHL1 (Fig. 1a and Extended Data Fig. 2a–c), the founding member of the FHL protein family[11]. FHL1 is characterized by four-and-a-half highly conserved LIM domains with two zinc fingers arranged in tandem[11]. FHL1 is highly expressed in skeletal muscles and heart[3,11]. There are three human FHL1 splice variants: FHL1A, FHL1B and FHL1C[3,12,13]. FHL1A is most-abundantly expressed in striated muscles[3] and fibroblasts[14]. The two other variants, FHL1B and FHL1C, are expressed in muscles, brain and testis[12,13]. We validated the requirement of FHL1 in CHIKV21 infection with two distinct guide (g)RNAs targeting the three *FHL1* isoforms (Extended Data Fig. 2a). We generated HAP1 and HEK293T *FHL1*-knockout clones (ΔFHL1)

and confirmed gene editing (Extended Data Fig. 2d–f). *FHL1* knockout did not alter cell proliferation and viability (Extended Data Fig. 2g). CHIKV21 infection and release of infectious particles was markedly inhibited in ΔFHL1 cells (Fig. 1b and Extended Data Fig. 3a–d). Trans-complementation of ΔFHL1 cells with a human cDNA encoding FHL1A, but not FHL1B or FHL1C, restored both the susceptibility to CHIKV21 and release of the virus (Fig. 1c and Extended Data Fig. 4a, b). Expression of FHL2, a member of the FHL family that is expressed in the heart[15], restored CHIKV infection in ΔFHL1 cells, albeit with a lower efficiency than FHL1 (Extended Data Fig. 4c). FHL1 is important for infection by CHIKV strains belonging to the Asian (strain St Martin H20235 2013), the east, central, and south African strains Ross and Brazza (MRS1 2011) and the Indian Ocean (strain M-899) lineages (Fig. 1d). Notably, the requirement for FHL1 was less pronounced with CHIKV 37997, a strain from the West African genotype (Fig. 1d). We next tested the requirement of FHL1 for infection by other alphaviruses. O'nyong-nyong virus (ONNV), an Old World alphavirus that is phylogenetically close to CHIKV, showed a markedly reduced infection level in ΔFHL1 cells (Fig. 1e and Extended Data Fig. 3e). By contrast, other alphaviruses—such as Mayaro virus (MAYV), Sindbis virus (SINV), Semliki Forest Virus (SFV), Ross River virus, eastern equine encephalitis virus, western equine encephalitis virus and Venezuelan equine encephalitis virus—infected HAP1 cells in a FHL1-independent manner (Fig. 1e, f and Extended Data Fig. 3e). FHL1 is not necessary for infection by the flaviviruses dengue virus (DENV) or Zika virus (ZIKV) (Fig. 1g and Extended Data Fig. 3f). BeWo or HepG2 cells, which are poorly susceptible to CHIKV infection[16,17] and lack FHL1 (Extended Data Fig. 5a), became permissive to infection after expression of FHL1A (Fig. 1h and Extended Data Fig. 5b–d). This highlights the major role of FHL1A in the permissiveness of human cells to CHIKV.

To determine which step in the CHIKV life cycle requires FHL1, we challenged parental and ΔFHL1 cells with CHIKV and quantified the viral RNA at different time points (Fig. 2a). We did not observe any significant difference in CHIKV RNA levels in ΔFHL1 cells compared to wild-type cells at 2 h after infection (Fig. 2a). By contrast, a large reduction in CHIKV RNA was observed in ΔFHL1 cells as early as 6 h after infection (Fig. 2a) and this reduction was greater at 24 h after infection. We bypassed virus entry and uncoating by transfecting CHIKV RNA into controls or ΔFHL1 cells in the presence of NH$_4$Cl to inhibit further rounds of infection[8]. Viral replication was markedly impaired in ΔFHL1 cells compared to wild-type cells

[1]INSERM U944, CNRS UMR 7212, Genomes & Cell Biology of Disease Unit, Institut de Recherche Saint-Louis, Université de Paris, Hôpital Saint-Louis, Paris, France. [2]Institut Pasteur, Inserm U1117, Biology of Infection Unit, Paris, France. [3]Institute of Virology, University of Bonn Medical Centre, Bonn, Germany. [4]Institut Pasteur, G5 Evolutionary Genomics of RNA Viruses, Paris, France. [5]INSERM U1259 MAVIVH et Plateforme IBiSA de Microscopie Electronique, Université de Tours et CHRU de Tours, Tours, France. [6]Institut de Génétique Humaine, Laboratoire de Virologie Moléculaire, CNRS-Université de Montpellier, Montpellier, France. [7]Unité des Virus Émergents, Aix-Marseille Univ, IRD190, Inserm 1207, EFS-IRBA, Marseille, France. [8]Equipe Chimie & Biologie, Modélisation et Immunologie pour la Thérapie, Université Paris Descartes, CNRS UMR 8601, Paris, France. [9]Center of Research in Myology, Sorbonne Université, INSERM UMRS974, Paris, France. [10]Département de Neurologie, Centre Hospitalier Universitaire de Montpellier, Montpellier, France. [11]Laboratoire de Virologie et Département des Maladies Infectieuses, Hôpital Saint-Louis, APHP, Paris, France. [12]Université de Paris, Department of Infectious Diseases and Tropical Medicine, Necker-Enfants Malades University Hospital, APHP, Institut Imagine, Paris, France. [13]These authors contributed equally: Laurent Meertens, Ali Amara. *e-mail: laurent.meertens@inserm.fr; ali.amara@inserm.fr
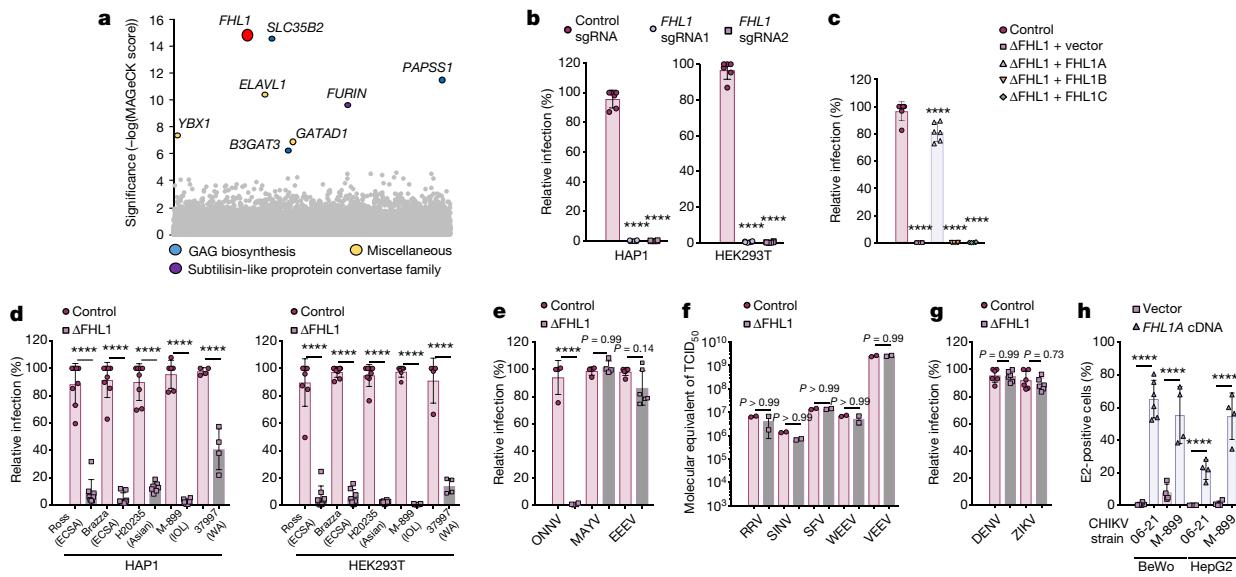
**Fig. 1 | FHL1 is important for infection by CHIKV and ONNV.**
**a**, Results of the CHIKV screen analysed by MAGeCK. Each circle represents an individual gene. The y axis shows the significance of sgRNA enrichment of genes in the selected population compared to the unselected control population. The x axis represents a random distribution of the genes. All genes with a false-discovery rate-adjust $P < 0.05$ are coloured (Benjamini-Hochberg procedure). **b**, E2 protein expression in control or ΔFHL1 cells infected with the CHIKV21 strain (multiplicity of infection (MOI) of 10). **c**, ΔFHL1 HAP1 cells were complemented with FHL1A, FHL1B or FHL1C isoforms, infected with CHIKV21 (MOI of 10) and stained for E2 protein expression at 48 h after infection. **b**, **c**, Data are mean ± s.d. $n = 3$ independent experiments performed in duplicate. One-way analysis of variance (ANOVA) with Dunnett's multiple comparison test. **d**, ΔFHL1 and control cells were inoculated with CHIKV Ross (MOI of 10), CHIKV Brazza (MOI of 10), CHIKV 20235 (MOI of 10), CHIKV-M (M-899) (MOI of 10) or CHIKV 37997 (MOI of 10) and analysed at 24 (HEK293T cells) or 48 h (HAP1 cells) after infection for E2 expression. ECSA, east, central, and south African strains; IOL, Indian Ocean strain; WA, west African strain. Data are mean ± s.d. $n = 4$ independent experiments performed in duplicate, except for CHIKV 37997, for which $n = 2$ independent experiments were performed in duplicate. One-way ANOVA with Tukey's multiple comparison test. **e**–**g**, ΔFHL1 and control HAP1 cells were inoculated with ONNV (MOI of 2), MAYV (MOI of 50), eastern equine encephalitis virus (EEEV) (MOI of 2), SINV, SFV, Ross

River virus (RRV), western equine encephalitis virus (WEEV), Venezuelan equine encephalitis virus (VEEV), DENV (MOI of 0.4) or ZIKV (MOI of 50). **e**, Infection was quantified 48 h after infection by flow cytometry using the anti-E2 3E4, 265 CHIKV monoclonal antibody or the anti-EEEV monoclonal antibody 1A4B6. $n = 2$ independent experiments performed in duplicate. One-way ANOVA with Tukey's multiple comparison test. **f**, Virus growth was assessed on day 4 after infection using RT–qPCR. Serial dilutions of infected supernatants titrated using the 50% tissue culture infectious dose ($TCID_{50}$) were used as quantification standards for RT–qPCR. Accordingly, results were expressed for each virus as 'molecular equivalents of $TCID_{50}$'. Data are representative of two independent experiments performed in duplicates. **g**, DENV or ZIKV infection was assessed by flow cytometry 48 h after infection using the anti-E protein 4G2 monoclonal antibody. $n = 3$ independent experiments performed in duplicate. **e**–**g**, Data are mean ± s.d. One-way ANOVA with Tukey's multiple comparison test. **h**, BeWo and HepG2 cells were transduced with FHL1A or a control vector and challenged with CHIKV21 (MOI of 5) or CHIKV M-899 (MOI of 2). Infection was quantified 2 days later by flow cytometry using the 3E4 monoclonal antibody. Data are mean ± s.d. $n = 2$ independent experiments performed in duplicate, except for BeWo cells infected with CHIKV21, for which $n = 3$ independent experiments were performed in duplicate. One-way ANOVA with Tukey's multiple comparison test. ****$P < 0.0001$.

(Fig. 2b and Extended Data Fig. 6a). To evaluate the contribution of FHL1 in viral RNA translation versus replication, we generated a replication-deficient CHIKV molecular clone (the GDD motif of the polymerase nsP4 was mutated to GAA), which encoded a *Renilla* luciferase (Rluc) fused to nsP3. Transfection of CHIKV(GAA) RNA in ΔFHL1 or control cells resulted in similar Rluc activities (Fig. 2c), indicating that FHL1 is dispensable for viral RNA translation. When similar experiments were performed with wild-type CHIKV RNA, a large increase in Rluc activity was observed in control—but not ΔFHL1—cells 24 h after infection (Fig. 2d), demonstrating that FHL1 is essential for viral RNA replication. Quantitative reverse-transcription PCR (RT–qPCR) experiments showed that ablation of FHL1 resulted in severely reduced synthesis of CHIKV negative-strand RNA (Fig. 2e). We investigated the effect of FHL1 in the production of dsRNA intermediates, which are a marker of viral replication complex assembly[18]. At 6 h after infection, a massive reduction in dsRNA-containing complexes was observed in ΔFHL1 cells stained with an anti-dsRNA monoclonal antibody compared to parental cells (Fig. 2f). Transmission electron microscopy analyses showed that the formation of plasma-membrane-associated spherules and cytoplasmic vacuolar membrane structures, which are alphavirus-induced platforms that are required for viral RNA synthesis[19], are absent in ΔFHL1 cells (Fig. 2g).

Confocal microscopy imaging showed that FHL1 displays a diffuse cytoplasmic distribution in uninfected human fibroblasts. In cells infected for 6 h, FHL1-containing foci appeared and colocalized with nsP3 (Extended Data Fig. 6b), a CHIKV non-structural protein that organizes viral replication in the cytoplasm[20,21]. CHIKV nsP3 contains a large C-terminal hypervariable domain (HVD)[20] that is known to mediate assembly of protein complexes and regulate RNA amplification[20,21]. Notably, FHL1 and FHL2 have been reported as putative nsP3(HVD)-binding partners in mass spectrometry analyses[21,22]. We experimentally validated the interaction between FHL1 and nsP3 (Fig. 2h, i and Extended Data Fig. 6c–g) and found that endogenous FHL1 co-immunoprecipitates with nsP3 from CHIKV-infected cells (Fig. 2h). Both FHL1A and FHL2 co-precipitated with CHIKV nsP3 (Extended Data Fig. 6d). The interaction between FHL1A and nsP3 is specific to CHIKV, as it was not observed with nsP3 from SINV or SFV (Extended Data Fig. 6e). In ΔFHL1 cells, nsP3 retained its ability to bind to G3BP1 and G3BP2, two components of the stress granules that have been implicated in CHIKV replication[21,23] (Extended Data Fig. 6e). We generated chimeric proteins in which the HVD region of CHIKV nsP3 is swapped with the corresponding domain of SINV nsP3 and vice versa. Whereas the CHIKV–SINV(HVD) chimeric protein lost its ability to bind to FHL1, the HVD of CHIKV in the context of the SINV nsP3 protein conferred binding to FHL1 (Extended Data Fig. 6f).
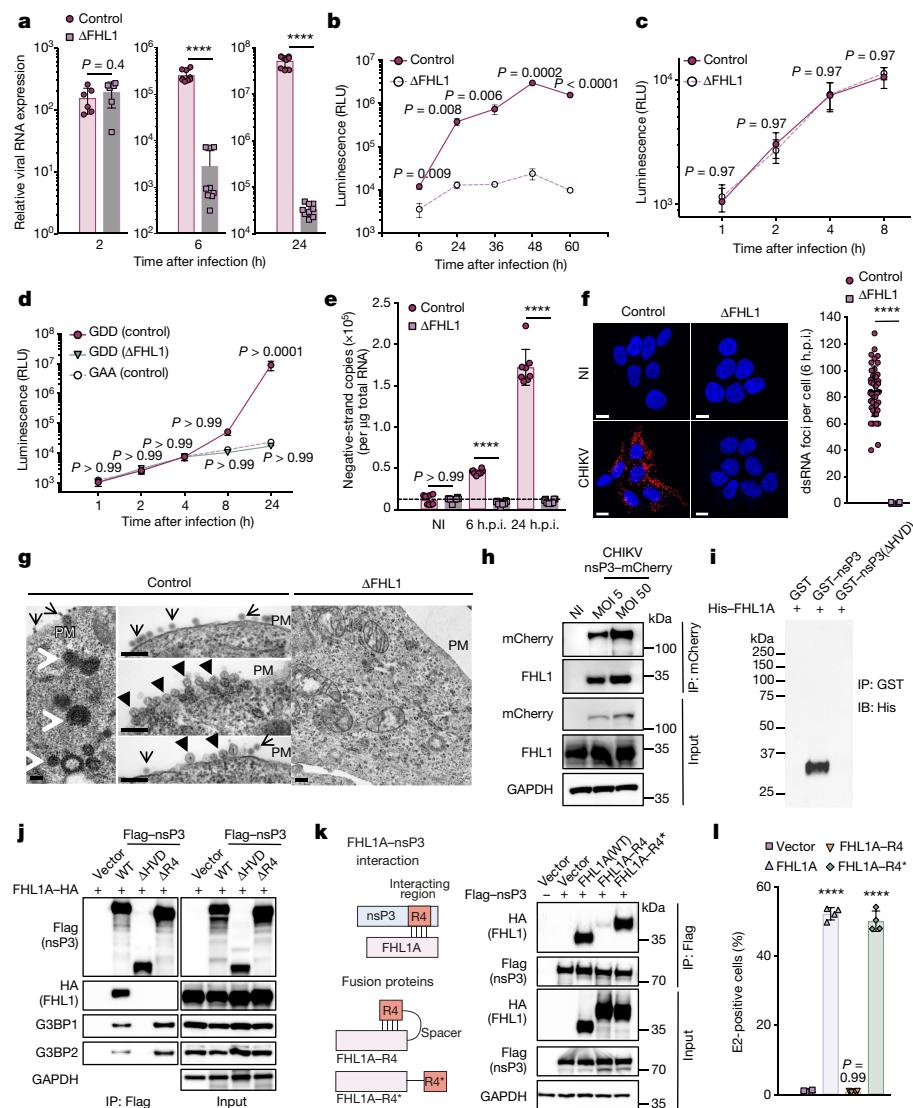
**Fig. 2 | FHL1 interacts with CHIKV nsP3 and is required for CHIKV RNA replication. a,** Control and ΔFHL1 HAP1 cells were inoculated with CHIKV21 (MOI of 10). At the indicated time points, cells were treated with trypsin to remove cell-surface-bound virus and viral RNA was quantified by RT–qPCR. Data are mean ± s.d. $n = 3$ independent experiments performed in triplicate, except for 2 h after infection, for which $n = 2$ independent experiments were performed. Two-tailed Student's $t$-test. **b,** Control or ΔFHL1 HEK293T cells were transfected with in vitro transcribed CHIKV-M RNA, expressing Gaussia luciferase (Gluc). Gluc activity was monitored at the indicated time points. RLU, relative light units. Data are mean ± s.e.m. $n = 3$ independent experiments performed in quadruplicate. Two-tailed multiple $t$-tests with Holm–Sidak correction. **c,** Control or ΔFHL1 HEK293T cells were transfected with a replication-deficient mutant CHIKV (CHIKV(GAA)) RNA, expressing Rluc and Rluc activity was monitored at the indicated time points. Data are mean ± s.e.m. $n = 3$ independent experiments performed in quadruplicate. Two-tailed multiple $t$-tests with Holm-Sidak correction). **d,** Control or ΔFHL1 HEK293T cells were transfected with a replication-competent (CHIKV(GDD)) or a replication-deficient mutant CHIKV (CHIKV(GAA)) capped RNA expressing Rluc. The Rluc activity was monitored at described in **c.** Data are mean ± s.e.m. $n = 3$ independent experiments performed in quadruplicate. Two-way ANOVA with Tukey's multiple comparison test. **e,** Negative-stranded viral RNA quantification by RT–qPCR from samples collected in **a.** h.p.i., hours post-infection; NI, not infected. Data are mean ± s.d. $n = 2$ independent experiments in quadruplicate. One-way ANOVA with a Tukey's multiple comparison test. Dashed line represents the experimental background threshold. **f,** Control and ΔFHL1 HEK293T cells were inoculated with CHIKV21 (MOI of 50). Left, representative images of infected cells stained with anti-dsRNA monoclonal antibody 6 h after infection. Scale bars, 10 μm. Right, number of foci per cell was quantified using the Icy software. Data are from 2 experiments, $n = 42$ cells in control and $n = 45$ cells in ΔFHL1 cells. Data are mean ± s.d. Two-tailed Student's $t$-test. **g,**

Transmission electron microscopy of control and ΔFHL1 HAP1 cells challenged with CHIKV21 (MOI of 100) at 24 h after infection. Data are representative of two independent experiments. Left, CPV-II structures containing attached nucleocapsids at their cytoplasmic side (white arrows) as well as viral particles at the cell surface (thin black arrows). Middle, replication spherules (arrowheads) together with viral particles (thin black arrows) at the plasma membrane (PM). Scale bars, 200 nm. **h,** Co-immunoprecipitation of endogenous FHL1 and CHIKV nsP3 from cell lysates of HEK293T cells infected with a CHIKV nsP3-mCherry reporter virus at MOI of 5 or 50. **i,** In vitro co-immunoprecipitation (IP) analysis of the direct interaction between CHIKV nsP3 and FHL1A through the HVD domain. Glutathione *S*-transferase (GST) precipitation of wild-type GST–nsP3 and GST–nsP3(ΔHVD) and immunoblot (IB) analysis of 6×His–FHL1A. **j,** HEK293T cells were co-transfected with plasmids encoding haemagglutinin (HA)-tagged FHL1A (FHL1A–HA) and Flag-tagged wild-type CHIKV nsP3, CHIKV nsP3(ΔHVD) or CHIKV lacking the amino acid region 423–454 (ΔR4). Proteins from cell lysates were immunoprecipitated with anti-Flag beads followed by immunoblot analysis with Flag, HA and G3BP1 and G3BP2 antibodies. **k,** Left, schematic representation of FHL1A fused to the nsP3 interacting region (FHL1A–R4) or a similar randomized sequence (FHL1A–R4*). Right, immunoassay of the interaction between CHIKV nsP3 and FHL1A fusion proteins in HEK293T cells co-transfected with Flag-tagged CHIKV nsP3 and HA-tagged FHL1A, FHL1A–R4 or FHL1A–R4* constructs. Proteins from cell lysates were immunoprecipitated with Flag antibody followed by immunoblot analysis with Flag and HA antibodies. **g–k,** Data are representative of three independent experiments. **l,** ΔFHL1 HEK293T cells were transfected with an empty vector or plasmids encoding FHL1A, FHL1A–R4 or FHL1A–R4*. Cells were incubated with CHIKV21 (MOI of 5) and infection was quantified 24 h after infection by flow cytometry. Data are mean ± s.d. $n = 2$ independent experiments performed in duplicate. One-way ANOVA with Dunnett's multiple comparison test. ****$P < 0.0001$.

**Fig. 3 | Primary cells from patients with FHL1 deficiency are resistant to CHIKV infection. a**, FHL1 expression in primary myoblasts and fibroblasts from healthy donors or patients with EDMD. CF1, CF2, control fibroblasts from healthy donors; CM1, CM2, control myoblasts from healthy donors, PF2, PF4, fibroblasts from patients with EDMD; PM1–PM3, myoblasts from patients with EDMD. Data are representative of two independent experiments. **b**, Cells from healthy donors (control) or patients with EMDM were inoculated with CHIKV expressing nsP3–mCherry. At 48 h after infection, cells were fixed and images were taken using a fluorescence microscope. Images are representative of three experiments. **c**, E2 protein expression in primary cells from healthy donors (control) or patients with EDMD infected with CHIKV21 (MOI of 2). Data are mean ± s.d. $n = 2$ independent experiments performed in duplicate for myoblasts; $n = 3$ independent experiments performed in duplicate for fibroblasts. One-way ANOVA with Tukey's multiple comparison test. **d**, Quantification of viral particles released in the supernatant of infected cells collected at 24, 48 and 72 h after infection. FIU, flow cytometry infectious units. Data are mean ± s.e.m. 2 independent experiments performed in duplicates for myoblasts; $n = 3$ independent experiments performed in duplicate for fibroblasts. Two-tailed multiple $t$-test with Holm–Sidak correction. **e**, Primary fibroblasts from a control (CF1) or from two patients with EDMD (PF2, PF4) were inoculated with CHIKV Ross, CHIKV Brazza, CHIKV H20235 strains or MAYV (MOI of 2) and analysed for E2 expression. Data are mean ± s.d. $n = 3$ independent experiments performed in duplicate. One-way ANOVA with Dunnett's multiple comparison test. **f, g**, Fibroblasts from control (CF1) or patients with EDMD (PF2, PF4) were transduced with a lentiviral vector encoding FHL1A or a control vector and then challenged with CHIKV21 (MOI of 2). **f**, Infection was quantified as described in **c**. Data are mean ± s.d. $n = 2$ independent experiments performed in duplicate. One-way ANOVA with Tukey's multiple comparison test. **g**, Supernatants were collected from infected cells at the indicated time points and viral titres were measured on VeroE6 cells. Data are mean ± s.e.m. $n = 2$ independent experiments performed in duplicate. Two-way ANOVA with Dunnett's multiple comparison test. ****$P < 0.0001$.

Pull-down experiments showed that FHL1A directly binds to wild-type nsP3 but not to the HVD-deficient variant (Fig. 2i and Extended Data Fig. 6g). We then mapped the binding region within CHIKV nsP3(HVD) that is responsible for interaction with FHL1A (Fig. 2j and Extended Data Fig. 7). The FHL1-binding domain, referred as HVD R4, is present in all CHIKV and ONNV strains, located upstream of the G3BP1- and G3BP2-binding sites[21] (Fig. 2j and Extended Data Fig. 7a). Deletion of the HVD R4 region strongly impaired the interaction of FHL1 with nsP3, without affecting G3BP1 and/or G3BP2 binding to the viral protein (Fig. 2j and Extended Data Fig. 7b). We generated

two chimeric FHL1A proteins that were fused either to the HVD R4 peptide (FHL1A–R4) or to a randomized peptide sequence of HVD R4 (FHL1A–R4*) as a positive control (Fig. 2k and Extended Data Fig. 7c), and assessed their ability to bind to nsP3. Whereas FHL1A-R4 failed to bind to nsP3 (Fig. 2k), FHL1A–R4* interacted with nsP3 as efficiently as wild-type FHL1A (Fig. 2k), indicating that the fused HVD R4 peptide probably hides the binding site of FHL1A to nsP3, inhibiting their interaction. Trans-complementation of ΔFHL1 cells with a cDNA encoding FHL1A–R4 did not restore CHIKV21 infection compared to
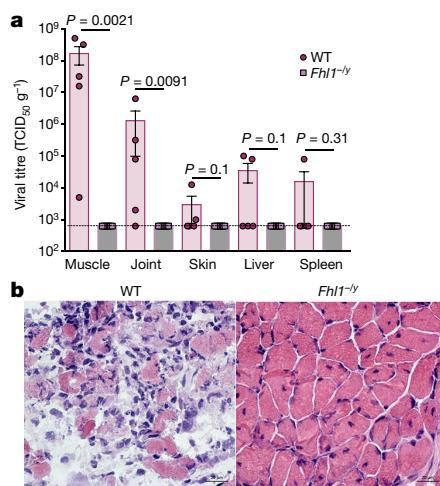


**Fig. 4 | FHL1 is a factor of susceptibility to CHIKV infection in mice. a**, Viral titres in tissues of 9-day-old mice. Wild-type (WT) male littermates ($n = 5$) and $Fhl1^{-/y}$ mice ($n = 7$) were inoculated with $10^5$ plaque-forming units of CHIKV by intradermal injection and euthanized 7 days after infection. The amount of infectious virus in tissues was quantified as the $TCID_{50}$. The dashed line indicates the detection
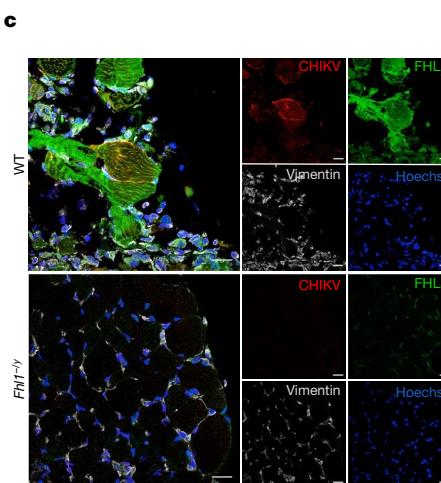
threshold. Data are mean ± s.e.m. Two-tailed $t$-test. **b**, Haematoxylin and eosin staining of transversal sections of skeletal muscle of CHIKV-inoculated mice. **c**, Immunostaining of nuclei (Hoechst), FHL1, vimentin and CHIKV in skeletal muscle of CHIKV-inoculated mice. **b, c**, Data are representative of $n = 3$ independent experiments.

FHL1A–R4* or wild-type FHL1A (Fig. 2l). In vitro transcribed RNA from a CHIKV molecular clone with a mutation in the FHL1-binding site (ΔR4 or R4*) showed a strong defect in replication in transfected HEK293T cells (Extended Data Fig. 7d). Together, these data show that the interaction between nsP3(HVD) and FHL1 is critical for the proviral function of FHL1.

Mutations in *FHL1* have been associated with X-linked myopathies[4,24], including Emery–Dreifuss muscular dystrophy (EDMD)[5], a rare genetic disorder characterized by early joint contractures, muscular wasting and adult-onset cardiac disease[25]. We studied the permissiveness of dermal fibroblasts and myoblasts to CHIKV that were obtained from four male patients with EDMD who carried mutations in *FHL1* as well as from two healthy donors (Extended Data Fig. 8a, b). FHL1 expression is severely reduced in primary cells from all four patients with EDMD (Fig. 3a). Infection studies showed that fibroblasts and myoblasts from patients with EDMD are resistant to CHIKV infection (Fig. 3b–e and Extended Data Fig. 8c, d) and exhibit a marked defect in the release of infectious particles (Fig. 3d). Cells of patients with EDMD remained highly susceptible to MAYV, which does not rely on FHL1 for replication (Fig. 3e). Trans-complementation of fibroblasts derived from patients with EDMD with a lentivirus that encodes FHL1A restored CHIKV viral antigen synthesis (Fig. 3f and Extended Data Fig. 8e) and the release of infectious particles (Fig. 3g).

To directly assess the role of FHL1 in the pathogenesis of CHIKV, we experimentally infected mice that did or did not express FHL1. Human and mouse FHL1 orthologues are highly conserved (Extended Data Fig. 9a). Mouse FHL1 interacts with CHIKV nsP3 and enhances viral infection, albeit less efficiently than its human orthologue (Extended Data Fig. 9b–d). Moreover, CHIKV infection is strongly impaired in the mouse muscle cell line C2C12 in which *Fhl1* is edited (Extended Data Fig. 9e, f). Susceptibility to CHIKV infection was tested in young male mice that were deficient in FHL1 expression and in young wild-type littermate male mice. CHIKV actively replicated in tissues of wild-type littermates as previously reported[16], but no infectious particles were detected in tissues of *Fhl1*$^{-/y}$ mice (Fig. 4a). Moreover, necrotizing myositis with large infiltrates and necrosis of the muscle fibres were observed in the skeletal muscle of wild-type littermates, whereas skeletal muscle of *Fhl1*$^{-/y}$ mice showed no detectable pathology (Fig. 4b). Immunolabelling against CHIKV E2 protein, FHL1 and vimentin in muscles revealed that in young wild-type mice, CHIKV mainly targets muscle fibres that express FHL1, whereas muscle cells of *Fhl1*$^{-/y}$ mice showed no labelling for CHIKV nor for FHL1 (Fig. 4c). These experiments demonstrate that *Fhl1*$^{-/y}$ mice are resistant to CHIKV infection.

In conclusion, our study identifies FHL1 as a critical host factor for CHIKV infection and pathogenesis. Indeed, the expression pattern of FHL1 reflects CHIKV tissue tropism. In addition to its direct implication for viral replication, hijacking of FHL1 by CHIKV may lead to cellular dysfunctions that contribute to the muscular and joint pains that are the hallmark of CHIKV disease. FHL1 interacts with CHIKV nsP3 to promote viral RNA synthesis. Deciphering the underlying mechanisms and understanding FHL1 selectivity for CHIKV will be essential to fully understand its pathogenesis and to develop novel therapeutic strategies to combat CHIKV disease.

## Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-019-1578-4.

1. Burt, F. J. et al. Chikungunya virus: an update on the biology and pathogenesis of this emerging pathogen. *Lancet Infect. Dis.* **17**, e107–e117 (2017).
2. Silva, L. A. & Dermody, T. S. Chikungunya virus: epidemiology, replication, disease mechanisms, and prospective intervention strategies. *J. Clin. Invest.* **127**, 737–749 (2017).
3. Greene, W. K., Baker, E., Rabbitts, T. H. & Kees, U. R. Genomic structure, tissue expression and chromosomal location of the LIM-only gene, *SLIM1*. *Gene* **232**, 203–207 (1999).
4. Schessl, J., Feldkirchner, S., Kubny, C. & Schoser, B. Reducing body myopathy and other FHL1-related muscular disorders. *Semin. Pediatr. Neurol.* **18**, 257–263 (2011).
5. Gueneau, L. et al. Mutations of the *FHL1* gene cause Emery–Dreifuss muscular dystrophy. *Am. J. Hum. Genet.* **85**, 338–353 (2009).
6. Ooi, Y. S., Stiles, K. M., Liu, C. Y., Taylor, G. M. & Kielian, M. Genome-wide RNAi screen identifies novel host proteins required for alphavirus entry. *PLoS Pathog.* **9**, e1003835 (2013).
7. Karlas, A. et al. A human genome-wide loss-of-function screen identifies effective chikungunya antiviral drugs. *Nat. Commun.* **7**, 11320 (2016).
8. Zhang, R. et al. Mxra8 is a receptor for multiple arthritogenic alphaviruses. *Nature* **557**, 570–574 (2018).
9. Tanaka, A. et al. Genome-wide screening uncovers the significance of *N*-sulfation of heparan sulfate as a host cell factor for chikungunya virus infection. *J. Virol.* **91**, 1–22 (2017).
10. Schuffenecker, I. et al. Genome microevolution of chikungunya viruses causing the Indian Ocean outbreak. *PLoS Med.* **3**, e263 (2006).
11. Shathasivam, T., Kislinger, T. & Gramolini, A. O. Genes, proteins and complexes: the multifaceted nature of FHL family proteins in diverse tissues. *J. Cell. Mol. Med.* **14**, 2702–2720 (2010).
12. Brown, S. et al. Characterization of two isoforms of the skeletal muscle LIM protein 1, SLIM1. Localization of SLIM1 at focal adhesions and the isoform slimmer in the nucleus of myoblasts and cytoplasm of myotubes suggests distinct roles in the cytoskeleton and in nuclear–cytoplasmic communication. *J. Biol. Chem.* **274**, 27083–27091 (1999).
13. Krempler, A., Kollers, S., Fries, R. & Brenig, B. Isolation and characterization of a new FHL1 variant (FHL1C) from porcine skeletal muscle. *Cytogenet. Cell Genet.* **90**, 106–114 (2000).
14. Pen, A. E. et al. A novel single nucleotide splice site mutation in *FHL1* confirms an Emery–Dreifuss plus phenotype with pulmonary artery hypoplasia and facial dysmorphology. *Eur. J. Med. Genet.* **58**, 222–229 (2015).
15. Chan, K. K. et al. Molecular cloning and characterization of FHL2, a novel LIM domain protein preferentially expressed in human heart. *Gene* **210**, 345–350 (1998).
16. Couderc, T. et al. A mouse model for chikungunya: young age and inefficient type-I interferon signaling are risk factors for severe disease. *PLoS Pathog.* **4**, e29 (2008).
17. Roberts, G. C. et al. Evaluation of a range of mammalian and mosquito cell lines for use in Chikungunya virus research. *Sci. Rep.* **7**, 14641 (2017).
18. Kim, D. Y. et al. New World and Old World Alphaviruses have evolved to exploit different components of stress granules, FXR and G3BP proteins, for assembly of viral replication complexes. *PLoS Pathog.* **12**, e1005810 (2016).
19. Jose, J., Taylor, A. B. & Kuhn, R. J. Spatial and temporal analysis of alphavirus replication and assembly in mammalian and mosquito cells. *mBio* **8**, e02294-16 (2017).
20. Götte, B., Liu, L. & McInerney, G. M. The enigmatic alphavirus non-structural protein 3 (nsP3) revealing its secrets at last. *Viruses* **10**, 105 (2018).
21. Meshram, C. D. et al. Multiple host factors interact with the hypervariable domain of chikungunya virus nsP3 and determine viral replication in cell-specific mode. *J. Virol.* **92**, e00838-18 (2018).
22. Mutso, M. et al. Mutation of CD2AP and SH3KBP1 binding motif in alphavirus nsP3 hypervariable domain results in attenuated virus. *Viruses* **10**, 226 (2018).
23. Scholte, F. E. M. et al. Stress granule components G3BP1 and G3BP2 play a proviral role early in chikungunya virus replication. *J. Virol.* **89**, 4457–4469 (2015).
24. Schessl, J. et al. Proteomic identification of FHL1 as the protein mutated in human reducing body myopathy. *J. Clin. Invest.* **118**, 904–912 (2008).
25. Bonne, G., Leturcq, F. & Ben Yaou, R. in *GeneReviews* (eds Adam, M. P. et al.) (University of Washington, 1993).

## METHODS

**Cell culture.** HAP1 cells (Horizon Discovery), which are derived from near-haploid chronic myeloid leukaemia KBM7 cells, were cultured in IMDM supplemented with 10% fetal bovine serum (FBS), 1% penicillin–streptomycin and GlutaMAX (Thermo Fisher Scientific). HEK293FT (Thermo Fisher Scientific), HEK293T (ATCC), Vero E6 (ATCC), BHK-21 (ATCC), HepG2 (gift of O. Schwartz), primary myoblasts and primary fibroblasts were cultured in DMEM supplemented with 10% FBS, 1% penicillin–streptomycin, 1% GlutaMAX and 25 mM HEPES. Human placenta choriocarcinoma BeWo cells were cultured in DMEM supplemented with 5% FBS, 1% penicillin–streptomycin, 1% GlutaMAX and 25 mM HEPES. AP61 mosquito (*Aedes pseudoscutellaris*) cells (gift from P. Despres) were cultured at 28 °C in Leibovitz medium supplemented with 10% FBS, 1% penicillin–strepto-mycin, 1% glutamine, 1× non-essential amino acids, 1× Tryptose phosphate and 10 mM HEPES. All cell lines were cultured at 37 °C in 5% $CO_2$ with the exception of AP61 cells, which were maintained at 28 °C without $CO_2$. Cell lines from ATCC were authenticated by the provider. HepG2, BeWo and AP61 were not further authenticated. All cell lines were tested for mycoplasma contamination.

**Virus strains and culture.** CHIKV21 (strain 06-21), ZIKV (HD78788) (gifts from P. Despres), CHIKV West Africa (strain 37997, accession AY726732.1) and DENV serotype 2 (16681) were propagated in mosquito AP61 cell monolayers with limited cell passages. CHIKV Brazza-MRS1 2011, CHIKV Ross, CHIKV St Martin H20235 2013 Asian, Ross River virus (strain 528v), MAYV (strain TC 625), ONNV (strain Dakar 234), SINV (strain Egypt 339), EEEV (strain H178/99), VEEV (strain TV83 vaccine), western equine encephalitis virus (strain 47A), SFV (strain 1745) were obtained from the European Virus Archive (EVA) collection and propagated with limited passages on Vero E6 cells.

pCHIKV-M-Gluc (see 'Plasmid constructions') and pCHIKV-mCherry molecular clones were derived from pCHIKV-M899, which is constructed from a CHIKV (strain BNI-CHIKV_899) strain isolated from a patient during the Mauritius outbreak in 2006. To generate infectious virus from CHIKV molecular clones, capped viral RNAs were generated from the NotI-linearized CHIKV plasmids using a mMESSAGE mMACHINE SP6 or T7 Transcription Kit (Thermo Fischer Scientific) according to the manufacturer's instructions. Resulting RNAs were puri-fied by phenol:chloroform extraction and isopropanol precipitation, resuspended in water, aliquoted and stored at −80 °C until use. Subsequently 30 μg of purified RNAs were transfected in BHK21 cells with Lipofectamine 3000 reagent and super-natants were collected 72 h later were used for viral propagation on Vero E6 cells.

For all of the viral stocks used in flow cytometry experiments, viruses were puri-fied through a 20% sucrose cushion by ultracentrifugation at 80,000*g* for 2 h at 4 °C. Pellets were resuspended in HNE1X pH 7.4 (HEPES 5 mM, NaCl 150 mM, EDTA 0.1 mM), aliquoted and stored at −80 °C. Viral stock titres were determined on Vero E6 cells by plaque-forming assay and are expressed as plaque-forming units (PFU) per ml. Virus stocks were also determined by flow cytometry as previously described[26,27]. In brief, Vero E6 cells were incubated for 1 h with 100 μl of tenfold serial dilutions of viral stocks. The inoculum was then replaced with 500 μl of culture medium and the percentage of E2-expressing cells was quantified by flow cytometry at 8 h after infection. Viral titres were calculated using the following formula and expressed as FIU per ml: Titre = (average percentage of infection) × (number of cells in well) × (dilution factor)/(ml of inoculum added to cells).

**Reagents.** The following antibodies were used: anti-FHL1 monoclonal antibody (MAB5938, R&D Systems), anti-FHL1 rabbit antibody (NBP1-88745, Novus Biologicals), anti-vimentin antibody (ab24525, Abcam), anti-GAPDH monoclo-nal antibody (SC-47724, Santa Cruz Biotechnology), polyclonal rabbit anti-HA (3724, Cell Signaling Technology), anti-Flag M2 monoclonal antibody (F1804, Sigma), anti-RFP (6G6, Chromotek), anti-CHIKV E2 monoclonal antibody (3E4 and 3E4 conjugated to Cy3), anti-alphavirus E2 monoclonal antibody (CHIK-265; gift from M. Diamonds), anti-EEEV E1 monoclonal antibody (MAB8754, Sigma), anti-pan-flavivirus E protein monoclonal antibody (4G2), anti-dsRNA J2 monoclo-nal antibody (Scicons), Alexa Fluor 488-conjugated goat anti-rabbit IgG (A11034, Invitrogen), Alexa Fluor-647-conjugated goat anti-chicken IgG (ab150175, Abcam), Alexa Fluor 488-conjugated goat anti-mouse IgG (115-545-003, Jackson ImmunoResearch), Alexa Fluor 647-conjugated goat anti-mouse IgG (115-606-062, Jackson ImmunoResearch), peroxydase-conjugated donkey anti-rabbit IgG (711-035-152, Jackson ImmunoResearch) and anti-mouse/HRP (P0260, DAKO Cytomotion). Flag magnetic beads (M8823, Sigma), HA magnetic beads (88837, Thermo Fisher Scientific) and anti-RFP coupled to magnetic agarose beads (RFP-Trap MA, Chromotek) were used for immunoprecipitation experiments.

**CRISPR genetic screen.** The GeCKO v.2 human CRISPR pooled libraries (A and B) encompassing 123,411 different sgRNA targeting 19,050 genes and cloned in the plentiCRISPR v.2[28] were purchased from GenScript. Lentiviral production was prepared independently for each half-library in HEK293FT cells by co-transfecting sgRNA plasmids with psPAX2 (from N. Manel) and pCMV-VSV-G at a ratio of 4:3:1 with Lipofectamine 3000 (Thermo Fisher Scientific). Supernatants were col-lected 48 h after transfection, cleared by centrifugation (750*g* for 10 min), filtered using a 0.45-μm filter and purified through a 20% sucrose cushion by ultracen-trifugation (80,000*g* for 2 h at 4 °C). Pellets were resuspended in HNE1X pH 7.4, aliquoted and stored at −80 °C. HAP1 cells were transduced by spinoculation (750*g* for 2 h at 32 °C) with each CRISPR–sgRNA lentiviral library at a multiplicity of infection (MOI) of 0.3 and a coverage of 500× the sgRNA representation. Cells were selected with puromycin for 8 days and expanded. Then, 60 million cells from each library were pooled and inoculated with CHIKV21, a viral strain isolated dur-ing the 2005–2006 CHIKV outbreak in La Reunion Island[10]. Approximately 5 days after infection, cytopathic effects were detectable and surviving cells were collected 2 weeks later. Genomic DNA was extracted from selected cells or uninfected pooled cells using a QIAamp DNA column (Qiagen), and inserted gRNA sequences were amplified and sequenced using next-generation sequencing on an Illumina MiSeq (Plateforme MGX, Institut Génomique Fonctionelle). gRNA sequences were ana-lysed using the MAGeCK software[29]. Additionally, gRNA sequences were analysed using the RIGER software following previously published recommendations[30].

**FHL1 editing.** FHL1 was validated using two independent sgR-NAs targeting exon 3 and exon 4, which are common to all FHL1 iso-forms. sgRNA1 (5′-GAGGACTCCCCCAAGTGCAA-3′) and sgRNA2 (5′-GCAGTCAAACTTCTCCGCCA-3′) were cloned into the plasmid lentiC-RISPR v.2 according to the recommendations of members of the Zhang laboratory. HAP1 and HEK293FT cells were transiently transfected with the plasmid express-ing individual sgRNAs and selected with puromycin until all mock-transfected cells died (approximately 72 h). Transfected cells were used to ascertain gRNA-driven resistance to the cytopathic effects caused by CHIKV, and clonal cell lines were isolated by limiting dilution and assessed by immunoblot for FHL1 expression.

**Infection assay.** For infection quantification by flow cytometry analysis, cells were plated in 24-well plates. Cells were infected for 24 (HEK293T cells) or 48 h (HAP1 cells), trypsinized and fixed with 2% (v/v) paraformaldehyde (PFA) diluted in PBS for 15 min at room temperature. Cells were incubated for 30 min at 4 °C with 1 μg ml⁻¹ of the 3E4 anti-E2 monoclonal antibody (for CHIKV strains and ONNV), the CHIKV 265 anti-E2 monoclonal antibody (for MAYV), the anti-E1 monoclonal antibody (for EEEV) or anti-pan-flavivirus E 4G2 antibody (for DENV and ZIKV). Antibodies were diluted in permeabilization flow cytometry buffer (PBS supplemented with 5% FBS, 0.5% (w/v) saponin, 0.1% sodium azide). After washing, cells were incubated with 1 μg ml⁻¹ of Alexa Fluor 488- or 647-conjugated goat anti-mouse IgG diluted in permeabilization flow cytometry buffer for 30 min at 4 °C. Acquisition was performed on an Attune NxT Flow Cytometer (Thermo Fisher Scientific) and analysis was done by using FlowJo software (TreeStar). To assess the release of infectious viral particles during infection, cells were inoculated for 3 h with viruses, washed once and then maintained in culture medium over a 72-h period. At the indicated time points, supernatants were collected and kept at −80 °C. Vero E6 cells were incubated with tenfold serial dilutions of supernatant for 24 h and E2 expression was quantified by flow cytometry as described above.

For detection of infected cells by immunofluorescence, control and ΔFHL1 HAP1 cells were plated on Laboratory-Tek II CC2 8-well glass slides (Nunc). Cells were inoculated with CHIKV21 (MOI of 20) or CHIKV nsP3–mCherry (MOI of 20) for 48 h, washed three times with cold PBS and fixed with 4% (v/v) PFA diluted in PBS for 20 min at room temperature. CHIKV E2 protein was stained with the 3E4 monoclonal antibody at 5 μg ml⁻¹, followed by a secondary staining with 1 μg ml⁻¹ of Alexa 488-conjugated goat anti-mouse IgG. Both antibodies were diluted in PBS supplemented with 3% (w/v) BSA and 0.1% saponin. Slides were mounted with ProLong Gold antifade reagent containing DAPI for nuclear staining (Thermo Fisher Scientific).

For colocalization experiments, cells infected with CHIKV nsP3–mCherry (MOI of 20) were stained with 10 μg ml⁻¹ of the anti-FHL1 monoclonal antibody, followed by secondary antibody staining with 1 μg ml⁻¹ of Alexa 488-conjugated goat anti-mouse IgG.

For detection of dsRNA foci, control and ΔFHL1 HEK293T cells were plated on Laboratory-Tek II CC2 8-well glass slides (Nunc) and infected with CHIKV21 (MOI of 50) for 4 or 6 h. After fixation with 4% (v/v) PFA diluted in PBS, cells were stained with 5 μg ml⁻¹ of the anti-dsRNA monoclonal antibody, followed by a secondary staining with 1 μg ml⁻¹ of Alexa 488-conjugated goat anti-mouse IgG. Both antibodies were diluted in PBS supplemented with 3% (w/v) BSA and 0.1% Triton X-100. Of note, no dsRNA foci were detectable at 4 h after infection.

Fluorescence microscopy images were acquired using a LSM 800 confocal microscope (Zeiss).

**Plasmid constructions.** To generate the C-terminal HA-tagged FHL1 iso-forms, the cDNAs of *FHL1A* (NM_001449.4), *FHL1B* (XM_006724746.2) and *FHL1C* (NM_001159703.1) were purchased from Genscript. Coding sequences were amplified with a common FHL1 forward primer (5′-CC**G GAGAATTC**CGCCGCCATGGCGGAGAAGTTTGACTGCCACTACTGC-3′); and specific *FHL1A* reverse primer (5′-AATAGTTTA**GCGGCCGC**TCAAGCG TAATCTGGAACATCGTATGGGTATCCTCCAGCGGCCGACAGCTTTTTG GCACAGTCGGGACAATACACTTGCTCC-3′); or specific *FHL1B* and *FHL1C*

reverse primer (5′-AATAGTTTA**GCGGCCGC**TCAAGCGTAATCTGGAACA
TCGTATGGGTATCCTCCAGCGGCCGACGGAGCATTTTTTGCAGTGGA
AGCAGTAGTCGTGCC-3′) (segments hybridizing with the target sequence are
underlined; restriction endonuclease sites for cloning are highlighted in bold); and
cloned into a pLVX-IRES-ZsGreen1 vector (Takara). Using the same approach,
the coding sequence of mouse *Fhl1* (NM_001077362.2) was amplified with a
mouse *Fhl1* forward primer (5′-CCGGAGAATTCGCCGCCATGGCTTCTCA
AAGCACACTCAGGTCCCTCC-3′) and mouse *Fhl1* reverse primer (5′-AATAG
TTTAGCGGCCGCTCAAGCGTAATCTGGAACATCGTATGGGTATCCTC
CAGCGGCCGACAGCTTTTTGGCACAGTCAGGGCAATACACCGCTC-3′),
and cloned into a pLVX-IRES-ZsGreen1 vector. The C-terminal HA-tagged
human *FHL2* coding sequence was synthesized by Genscript and subcloned
into a pLVX-IRES-ZsGreen1 vector. The pCI-neo-3×Flag plasmids express-
ing CHIKV nsP3 and nsP4, SINV and SFV nsP3 proteins were previously
described[31]. The CHIKV nsP3(ΔHVD), ΔR1 to ΔR4 proteins were generated by
site-directed mutagenesis (QuickChange XL Site-Directed Mutagenesis
Kit, Agilent) using the following sets of primers: ΔHVD forward (5′-CGTA
AGTCCAAGGGAATATTGATGATCTTCCCAGGAGTCTGC-3′) and
ΔHVD reverse (5′-GCAGACTCCTGGGAAGATCATCAATATTCCC
TTGGACTTACG-3′); ΔR1 forward (5′-GTACCTGTCGCGCCGCCCAGAGAG
CTGTGTCCGGTCGTACAAGAAAC-3′) and ΔR1 reverse (5′-GTTTCTTGTA
CGACCGGACACAGCTCTCTGGGCGGCGCGACAGGTAC-3′); ΔR2 forward
(5′-GAAACAGCGGAGACGCGTGACAGTACCGCCACGGAACCGAATC-3′)
and ΔR2 reverse (5′-GATTCGGTTCCGTGGCGGTACTGTCA
CGCGTCTCCGCTGTTTC-3′); ΔR3 forward (5′-CTTCTTACCAGGAG
AAGTGTGATGACTTGACAGACAGC-3′) and ΔR3 reverse (5′-GCTGTCTGTC
AAGTCATCACACTTCTCCTGGTAAGAAG-3′); ΔR4 forward (5′-GACG
AGAGAGAGGGAATACACCGAGTACCGCCACGGAACCGAATC-3′)
and ΔR4 reverse (5′-GATTCGGTTCCGTGGCGGTACTCGGTGTTATATTCC
CTTCTCTCTCGTC-3′).

The plasmids expressing chimeric nsP3 CHIKV–HVD SINV and nsP3
SINV–HVD CHIKV were obtained as follows. First, the DNA sequence
encoding the N-terminal parts of the CHIKV or SINV nsP3 (MD–AUD
region) were obtained by PCR using the pCI-neo-3×Flag expression plas-
mids as templates and the following sets of primers: 3×Flag NotI for-
ward (5′-ACTGAGCGGCCGCATGGACTACAAAGACCATGAC-3′)
and overlap CHIKV–SINV reverse (5′-GCTGTTCTGGCACTTCTATAT
ATTCCCTTGGACTTACG-3′), or 3×Flag NotI forward and overlap SINV–CHIKV
reverse (5′-CAGACTCCTGGGAAGATCTGTACTTACGGGCGGGAAC-3′)
for CHIKV and SINV constructs, respectively. HVD coding sequences were also
generated by PCR using the following primers: overlap CHIKV–SINV forward
(5′-CGTAAGTCCAAGGGAATATATAGAAGTGCCAGAACAGC-3′) and nsP3
SINV BamHI reverse (5′-ACTGAGGATCCTTAGTATTCAGTCCTCCTGCTC-3′)
for SINV HVD; and overlap SINV–CHIKV forward (5′-GTTCCCGCCCGT
AAGTACAGATCTTCCCAGGAGTCTG-3′) and nsP3 CHIKV BamHI reverse
(5′-ACTGAGGATCCTCATAACTCGTCGTCCGTG-3′) for CHIKV HVD. Next,
the CHIKV–HVD–SINV and SINV–HVD–CHIKV PCR fragments were obtained
by overlap extension PCR using the previously obtained PCR products and the
following sets of primers: 3×Flag NotI forward and nsP3 SINV BamHI reverse,
or nsP3 CHIKV BamHI reverse. Finally, the chimeric PCR fragments were cloned
into a NotI–BamHI-digested pLVX-IRES-ZsGreen1 vector (Takara).

The plasmid expressing FHL1A–R4 and FHL1A–R4* fusion pro-
teins were also obtained by overlap extension PCR approach. First,
the *FHL1A* part, which is common to both constructs, was amplified
from a cDNA template (Genscript, NM_001449.4) using the common
*FHL1* forward primer (5′-CCGGAGAATTCGCCGCCATGGCGGA
GAAGTTTGACTGCCACTACTGC-3′) and the overlap *FHL1A* fusion
reverse primer (5′-CGCCCTGGAAGTACAGGTTCTCGCCGCCGCCC
AGCTTTTTGCACAGTCGGGACAATAC-3′). Second, nsP3 R4 and
R4* portions were obtained by PCR using either the pCI-neo-3×Flag-nsP3
expression plasmid or the pCHIKV-SG45-R4* plasmid (containing the ran-
domized R4 region) as templates and the following set of primers: overlap
*FHL1* fusion forward (5′-CGAGAACCTGTACTTCCAGGGCGGCGGCGG
CCCCATGGCTAGCGTCCGATTCTTTAG-3′) and *FHL1* fusion reverse
(5′-AATAGTTTAGCGGCCGCTCAAGCGTAATCTGGAACATCGTATGGGTA
GCCGCCGCCCGGTGGTGCCTGAAGAGACATTGCTG-3′) for the R4
construct; or *FHL1* fusion Random reverse primer (5′-AATAGTT
TAGCGGCCGCTCAAGCGTAATCTGGAACATCGTATGGGTAGCCGCC
GCCCCTCACCTCGGCGCACATGG-3′) for the randomized R4* construct.
Next, the FHL1A–R4 and FHL1A–R4* PCR fragments were obtained by PCR
using the previously obtained PCR products and the outer sets of primers: *FHL1A*
forward and *FHL1* fusion reverse or *FHL1* fusion Rand reverse. Amplification
fragments were cloned into a NotI–EcoRI-digested pLVX-IRES-ZsGreen1 vector
(Takara).

To obtain pCHIKV-M-Gluc, a viral sequence encompassing the CHIKV 26S pro-
moter and part of the capsid protein sequence was amplified from pCHIKV-M using
primers (5′-TATGCGTTTAAACCATGGCCACCTTTGCAAGCTCCAGATC-3′)
and (5′-GCTTCTTATTCTTCCGATTCCTGCGTGG-3′), cut with PmeI and
BssHII and assembled together with an AgeI–PmeI fragment from pCHIKVRepl-
Gluc[32] into an AgeI–BssHII cut vector. From the resulting plasmid the AgeI–
BssHII fragment was released and ligated together with a BssHII–SfiI fragment
from pCHIKV-M[33] into pCHIKV-M cut with AgeI and SfiI.

To establish pCHKV-Rluc-GAA, two PCR fragments were amplified from
wild-type CHIKV using primers CHIKV 5590 F (5′-AGACTTCTTACCAG
GAGAAGTG-3′) and Bo422 (5′-CGACTCCATGTATTATGTTACCCGCTGC
GATGAAGGCCGCGCACGCGG-3′) or Bo421 (5′-CCGCGTGCGCGGCCT
TCATCGCAGCGGGTAACATAATACATGGAGTCG-3′) and CHIKV 8512 R
(5′-GAAGTTGTCCTTGGTGCTGC-3′), respectively. The obtained fragments
were fused via PCR amplification using the outer primers CHIKV 5590 forward
and CHIKV 8512 reverse. The resulting fragment was cut with AgeI and BglI and
inserted into pCHIKV-Rluc cut with the same restriction enzymes.

For generation of CHIKV-Rluc-ΔR4 and CHIKV-Rluc-R4*, PCR fragments
encompassing the desired changes were first amplified and assembled as follows.
For CHIKV-Rluc-ΔR4, two fragments were amplified from CHIKV-Rluc using
Bo408 (5′-CACCACGTGCTCCTGGTCAGTG-3′) and Bo1259 (5′-GATTCG
GTTCCGTGGCGGTACTCGGTGTTATATTCCCTTCTCTCTCGTCA-3′) or
Bo1258 (5′-tGACGAGAGAGAAGGGAATATAACACCGAGTACCGCCACG-
GAACCGAATC-3′) and Bo409 (5′-GACTTCCTCCAGGGTGTTCACC-3′) and
Bo409 (5′-GACTTCCTCCAGGGTGTTCACC-3′), respectively, and were fused
together using the outer primers Bo408 and Bo409. For CHIKV-Rluc-R4*, the
randomized sequence cassette was obtained sequentially from three successive
PCRs. First, the PCR fragment was generated using primers Bo1260 (5′-AGCAC
CGTGCCCCTGCCCGCCCTGAGGAGGGCCAGCTTCGCCGACACCATGG
AGCAGACC-3′) and Bo1261 (5′-CCTCACCTCGGCGCACATGGGGAACTG
CTCGGCCACGGTCTGCTCCATGGTGTCGGCGAA-3′). Then, it was fused
at the 5′ end with a PCR fragment amplified from CHIKV-Rluc with Bo408 and
Bo1262 (5′-TCAGGGCGGGCAGGGGCACGGTGCTtgttatattcccttctctctcgtca-3′).
Next, the resulting fragment was further fused at the 3′ end with a PCR
fragment amplified from CHIKV-Rluc with Bo1263 (5′-GTTCCCC
ATGTGCGCCGAGGTGAGGccgagtaccgccacggaaccgaatc-3′) and Bo409, using the
outer primers Bo408 and Bo409. Finally, the PCR fragments containing the ΔR4
and R4* mutations were cut with SacII and AgeI and fused with a NgoMIV–SacII
fragment derived from CHIKV-Rluc (SG45) and cloned into a NgoMIV–AgeI-
digested SG45 plasmid.

**Trans-complementation and overexpression experiments.** The lentiviral plas-
mids containing FHL1 isoforms were packaged as described above (see 'CRISPR
genetic screen'). Cells of interest were stably transduced by spinoculation (750*g* for
2 h at 32 °C) with these lentiviruses and, when necessary, sorted for GFP-positive
cells by flow cytometry. For trans-complementation assays, cells were inoculated
with CHIKV21 for 48 h. Cells were then collected and processed for E2 expres-
sion by flow cytometry. For ectopic expression, cells were plated on 24-well plates
($5 \times 10^4$) and incubated with CHIKV-M-Gluc and CHIKV21, and either processed
for E2 expression by flow cytometry or infectious virus yield quantification on
Vero E6 cells.

**Kinetic of infection by RT–qPCR assay.** Control and ΔFHL1 HAP1 cells were
plated on 60-mm dishes (400,000 cells) and inoculated with CHIKV21 (MOI of
5). At the indicated time points, cells were washed three times with PBS, incu-
bated with 0.25% trypsin for 5 min at 37 °C to remove cell-surface-bound par-
ticles, and total RNA was extracted using the RNeasy Plus Mini kit (Qiagen)
according to the manufacturer's instructions. cDNAs were generated from 500 ng
total RNA using the Maxima First Strand Synthesis Kit following the manufac-
turer's instructions (Thermo Fisher Scientific). Amplification products were
incubated with 1 unit of RNase H for 20 min at 37 °C, followed by 10 min at
72 °C for enzyme inactivation, and diluted tenfold in DNase/RNase-free water.
RT–qPCR was performed using a Power Syber green PCR master Mix (Fisher
Thermo Scientific) on a Light Cycler 480 (Roche). The primers used for RT–
qPCR were: E1-C21 forward (5′-ACGCAGTTGAGCGAAGCAC-3′), E1-C21
reverse (5′-CTGAAGACATTGGCCCCAC-3′) for viral RNA quantification, and
Quantitect primers for *GAPDH* were purchased from Qiagen. Quantification using
relative expression was performed based on the comparative threshold cycle ($C_t$)
method, using *GAPDH* as endogenous reference control. CHIKV negative-strand
RNA was quantified as previously described[34]. In brief, cDNA was generated
from 1 μg total RNA using a primer containing a 5′ tag sequence CHIKV(−)
tag (5′-GGCAGTATCGTGAATTCGATGCCGGCTGTTACCGTCCCCATTCC-3′)
and the SuperScript II reverse transcriptase following the manufacturer's
instructions (Thermo Fisher Scientific). Amplifications products were diluted
tenfold and used for RT–qPCR with the following primers CHIKV(−) for-
ward (5′-GGCAGTATCGTGAATTCGATGC-3′) and CHIKV(−) reverse

(5′-ACTGCTGAGTCCAAAGTGGG-3′). The 133-bp sequence corresponding to the amplified cDNA was synthesized by Genescript and serially diluted (650 to $6.5 \times 10^9$ genes copies $\mu l^{-1}$) to generate standard curves.

**Genomic viral RNA transfection and kinetic of viral amplification.** To assess CHIKV RNA replication within the cells, we transfected control and ΔFHL1 cells with capped genomic viral RNA generated from pCHIKV-M-Gluc (see 'Virus strains and culture'). Cells were plated on 48 well plate ($3 \times 10^4$ cells) and transfected with 100 ng of purified RNA using the Lipofectamine MessengerMax reagent according to the manufacturer's instructions (Thermo Fisher Science), and cells were cultured in the absence or presence of 15 mM NH₄Cl to prevent subsequent viral propagation. At specific times, cells were washed once with PBS and lysed with *Gaussia* lysis buffer. Lysates were kept at −20 °C until all samples were collected. Luciferase activity was measured using the Pierce *Gaussia* Luciferase Glow assay kit on a TriStar2 LB 942 with 20 μl of cell lysate, 20 μl of substrate and 2 s integration time.

The same experimental approach was used to monitor luciferase activity from capped genomic viral RNA generated from wild-type pCHIKV-Rluc (SG45), pCHIKV-Rluc-GAA, pCHIKV-Rluc-ΔR4 and pCHIKV-Rluc-R4* mutants. Luciferase activity was measured using the *Renilla* luciferase assay system (Promega) on a TriStar2 LB 942 with 20 μl of cell lysate, 20 μl of substrate and 2.5 s integration time.

**Immunoblots.** Cell pellet were lysed in Pierce IP Lysis Buffer (Thermo Fisher Scientific) containing Halt protease and phosphatase inhibitor cocktails (Thermo Fischer Scientific) for 30 min at 4 °C. Equal amounts of protein, determined by DC Protein Assay (BioRad), were prepared in LDS sample buffer 4× (Pierce) containing 25 mM dithiothreitol (DTT) and heated at 95 °C for 5 min. Samples were separated on Bolt 4–12% Bis-Tris gels in Bolt MOPS SDS Running Buffer (Thermo Scientific), and proteins were transferred onto a PVDF membrane (BioRad) using the Power Blotter system (Thermo Fischer Scientific). Membranes were blocked with PBS containing 0.1% Tween-20 and 5% non-fat dry milk and incubated overnight at 4 °C with primary antibodies. Staining was revealed with corresponding horseradish peroxidase (HRP)-coupled secondary antibodies and developed using SuperSignal West Dura Extended Duration Substrate (Thermo Fisher Scientific) following the manufacturer's instructions. The signals were acquired through Fusion Fx camera (VILBERT Lourmat).

**Co-immunoprecipitation assay.** HEK293T cells were plated in 10-cm dishes ($5 \times 10^6$ cells per dish). After 24 h, the cells were transfected with a total of 15 μg DNA expression plasmids (7.5 μg of each plasmid in co-transfection assays). After 24 h of transfection, the cells were washed once with PBS and collected with a cell scraper. After centrifugation (400$g$ for 5 min), cell pellets were lysed for 30 min in cold immunoprecipitation lysis buffer supplemented with Halt protease and phosphatase inhibitor cocktails, and then cleared by centrifugation for 15 min at 6,000$g$. Supernatants were incubated overnight at 4 °C, with either anti-Flag magnetic beads or anti-HA magnetic beads (see 'Reagents'). Beads were washed three times with BO15 buffer (20 mM Tris-HCl pH 7.4, 150 mM NaCl, 5 mM MgCl₂, 10% glycerol, 0.5 mM EDTA, 0.05% Triton X-100, 0.1% Tween-20). The retained complexes were eluted twice with either 3×Flag peptide (200 μg ml⁻¹; Sigma-Aldrich F4799-4MG) or HA peptide (400 μg ml⁻¹; Roche 11666975001) for 30 min at room temperature. Samples were prepared and immunoblotted as described above. For input, 1% of whole-cell lysates was loaded on the gel.

**Bacterial expression, purification and GST pull-down assay.** To express nsP3 and nsP3(ΔHVD) as GST fusion proteins, their respective open reading frames were subcloned into pGEX-4T-1. Similarly, *FHL1A* cDNA was subcloned into pET47b(+) and expressed as a 6×His fusion protein. The following oligonucleotides were used to amplify nsP3 and nsP3ΔHVD cDNAs (sense, 5′-CCCCGGAATTCCATGGCACCGTCGTACCGGGTAA-3′; antisense, 5′-CCGCTCGAGTCATAACTCGTCGTCCGTGTCTG-3′) and *FHL1A* (sense, 5′-CCGGAATTCCATGGCGGAGAAGTTTGACTGCC-3′; antisense, 5′-CCGCTCGAGTTACAGCTTTTTGGCACAGTC-3′). *Escherichia coli* strain BL21 Star (Invitrogen) was transformed with recombinant expression vectors encoding GST−nsP3, GST−nsP3(ΔHVD) or 6×His−FHL1A recombinant proteins. Transformed bacteria were induced with isopropylthio-β-Dgalactoside (IPTG) for 3 h at 37 °C. Cells were collected by centrifugation and the pellets were resuspended in lysis buffer containing lysozyme (1 mg ml⁻¹), incubated for 30 min at 4 °C followed by three subsequent freeze–thaw cycles and sonication. The bacterial lysates were centrifuged at 13,000 r.p.m for 20 min and the supernatants were incubated with glutathione-sepharose beads for GST−nsP3 and GST−nsP3(ΔHVD) or on a Ni-NTA column (Qiagen) for 6×His−FHL1A. Column washing and recombinant protein elution were performed according to the manufacturer's instructions. In brief, 5 μl of eluted GST-fusion proteins and 3 μl of Ni-NTA-eluted 6×His−FHL1A were analysed by SDS−PAGE and proteins were visualized by Coomassie staining. For pull-down assays, GST-, GST−nsP3- or GST−nsP3(ΔHVD)-bound beads were incubated with 6×His−FHL1A for 1 h at 4 °C in the presence of 100 μM ZnSO₄. The resin was washed extensively with

a buffer containing 500 mM KCl. The beads were then resuspended in Laemmli buffer, resolved by SDS−PAGE and the presence of 6×His−FHL1A was assessed by western blot using the anti-FHL1 antibody.

**Genetic analysis, fibroblasts and myoblasts from patients with EDMD.** Dermal fibroblasts and myoblasts were taken from four patients carrying *FHL1* mutations. *FHL1* was analysed as previously reported[5], as the patients had, among other symptoms, features that were reminiscent of EDMD. Patients P1, P2 and P3 have been previously reported[5] with detailed clinical descriptions (respectively as patient F321-3, F997-8 and F1328-4), whereas information for patient P4 was not previously published. In brief, patient P4 had myopathy with joint contractures, hypertrophic cardiomyopathy, vocal cord palsy, short stature, alopecia, skin abnormalities and facial dysmorphism. In this patient, *FHL1* analysis revealed an insertion of a full-length LINE-1 retrotransposon sequence together with poly(A) tail of unknown length (indicated by a '?' thereafter) after 27 bp of the start of exon 4 (c.183–184ins, *LINE1*; ?; 171–183) that results at the mRNA level in altered splicing with retention of 108 bp of the inserted LINE sequence leading to a predicted premature termination codon and shorter *FHL1A* (Extended Data Fig. 8b).

**Ethics statement.** All materials (skin and/or muscle biopsies) from patients and controls included in this study were taken with the informed consent of the donors and with approval of the local ethical boards (that is, P1, Tokyo Women's Medical University, Japan; P2, King Saudi University, Saudi Arabia; P3, University Hospital of Lille, France; P4, University Hospital of Montpellier, France). All procedures were followed alongside the usual molecular diagnostic procedure during patient follow-up, and in accordance with the ethical standards of the responsible national committee on human experimentation.

**In vivo studies.** Animals were housed in the Institut Pasteur animal facilities accredited by the French Ministry of Agriculture for performing experiments on live rodents. Work on animals was performed in compliance with French and European regulations on care and protection of laboratory animals (EC Directive 2010/63, French Law 2013-118, 6 February 2013). All experiments were approved by the Ethics Committee 89 (and registered under the reference APAFIS#6954-2016091410257906 v.2). FHL1-deficient male mice (*Fhl1⁻/ʸ*) or wild-type male littermates were obtained by crossing *Fhl1* heterozygous females[35] with wild-type male Black Swiss mice. Subsequently, 9-day-old *Fhl1⁻/ʸ* and wild-type male littermates, were injected with CHIKV21 ($10^5$ PFU per 20 μl) by intradermal route and viral load was determined in tissues on day 7 after inoculation. Virus titres in tissue samples were determined on Vero E6 cells as TCID₅₀ g⁻¹. For histology experiments, muscles were snap-frozen in isopentane cooled by liquid nitrogen for cryo-sectioning then processed for histological staining (haematoxylin and eosin) or immunolabelling.

**Transmission electron microscopy.** Cells were scraped and fixed for 24 h in 1% glutaraldehyde, 4% PFA (Sigma) in 0.1 M phosphate buffer (pH 7.2). Samples were then washed in PBS and post-fixed for 1 h by incubation with 2% osmium tetroxide (Agar Scientific). Cells were subsequently fully dehydrated in a graded series of ethanol solutions and propylene oxide. An impregnation step was performed with a mixture of (1:1) propylene oxide/Epon resin (Sigma) and then left overnight in pure resin. Samples were embedded in Epon resin (Sigma), which was allowed to polymerize for 48 h at 60 °C. Ultra-thin sections (90 nm) of these blocks were obtained with a Leica EM UC7 ultramicrotome. Sections were stained with 2% uranyl acetate (Agar Scientific), 5% lead citrate (Sigma) and observations were made with a transmission electron microscope (JEOL 1011).

**Cell viability assay.** Cell viability and proliferation were assessed using the CellTiter-Glo 2.0 assay (Promega) according to the manufacturer's protocol. In brief, cells were plated in 48-well plates ($3 \times 10^4$). At specific times, 100 μl of CellTiter-Glo reagent were added to each well. After a 10-min incubation, 200 μl from each well was transferred to an opaque 96-well plate (Cellstar, Greiner Bio-One) and luminescence was measured on a TriStar2 LB 942 (Berthold) with 0.1 s integration time.

**Statistical analysis.** Graphical representation and statistical analyses were performed using Prism 7 software (GraphPad software). Unless otherwise stated, results are shown as mean ± s.d. from at least two independent experiments in duplicates. Differences were tested for statistical significance using an unpaired two-tailed *t*-test, or one-way or two-way ANOVA with multiple comparison post hoc test.

**Reporting Summary.** Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

## Data availability

The authors declare that the data supporting the findings are available within the paper and its Supplementary Information. Source Data for Figs. 1–4 are provided with the paper. All other data are available from the corresponding authors upon request.

26. Medina, F. et al. Dengue virus: isolation, propagation, quantification, and storage. *Curr. Protoc. Microbiol.* **15**, 15D.2.1–15D.2.24 (2012).
27. Meertens, L. et al. The TIM and TAM families of phosphatidylserine receptors mediate dengue virus entry. *Cell Host Microbe* **12**, 544–557 (2012).
28. Shalem, O. et al. Genome-scale CRISPR–Cas9 knockout screening in human cells. *Science* **343**, 84–87 (2014).
29. Li, W. et al. MAGeCK enables robust identification of essential genes from genome-scale CRISPR/Cas9 knockout screens. *Genome Biol.* **15**, 554 (2014).
30. Joung, J. et al. Genome-scale CRISPR–Cas9 knockout and transcriptional activation screening. *Nat. Protoc.* **12**, 828–863 (2017).
31. Pellet, J. et al. ViralORFeome: an integrated database to generate a versatile collection of viral ORFs. *Nucleic Acids Res.* **38**, D371–D378 (2010).
32. Gläsker, S. et al. Virus replicon particle based chikungunya virus neutralization assay using *Gaussia* luciferase as readout. *Virol. J.* **10**, 235 (2013).
33. Kümmerer, B. M., Grywna, K., Gläsker, S., Wieseler, J. & Drosten, C. Construction of an infectious chikungunya virus cDNA clone and stable insertion of mCherry reporter genes at two different sites. *J. Gen. Virol.* **93**, 1991–1995 (2012).
34. Plaskon, N. E., Adelman, Z. N. & Myles, K. M. Accurate strand-specific quantification of viral RNA. *PLoS ONE* **4**, e7468 (2009).
35. Domenighetti, A. A. et al. Loss of FHL1 induces an age-dependent skeletal muscle myopathy associated with myofibrillar and intermyofibrillar disorganization in mice. *Hum. Mol. Genet.* **23**, 209–225 (2014).

**Author contributions** L.M. and A.A. conceived the study. L.M., M.L.H., T.C., V.K., A.B., L.B.-M., C. Delaugerre, M.L. and A.A. designed the experiments. L.M. performed the CRISPR–Cas9 screening and infection studies with L.B.-M. M.L.H. characterized the FHL1 and nsP3 interactions and performed the immunoprecipitation and western blot studies with help of V.K. A.L. validated the gRNA targeting human and mouse FHL1 and generated the *FHL1*-knockout cells described in this study and performed infection studies. A.B. performed the immunofluorescence microscopy experiments and some infection studies. L.M. and E.S.-L. analysed the gRNA sequencing studies and identified the hits. J.B.-G. and P.R. performed the electron microscopy experiments. C. Doyen and M.B. performed the GST pull-down experiments. B.M.K. generated the CHIKV molecular clones described in this study and P.-O.V. provided key CHIKV reagents. L.P. and X.d.L. provided the alphavirus strains and performed infection studies. T.C. and M.L. conceived and performed the in vivo studies and provided expertise in the design of CHIKV experiments. S.R. performed virus titration assays and mice genotyping. T.G. performed immunofluorescence experiments in mouse tissues with T.C. A.B.-L., R.B.Y., L.G., R.J.-M. and G.B. provided myoblasts and fibroblasts from patients with EDMD and the description of a new *FHL1* mutation in EDMD disease. L.M. and A.A. wrote the initial manuscript draft, M.L. edited the draft and the other authors contributed to editing it in its final form.

**Competing interests** The authors declare no competing interests.

**Additional information**
**Supplementary information** is available for this paper at https://doi.org/10.1038/s41586-019-1578-4.
**Correspondence and requests for materials** should be addressed to L.M. or A.A.
**Peer review information** *Nature* thanks Laurie A. Silva and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.
**Reprints and permissions information** is available at http://www.nature.com/reprints.

**a**

Transduction of HAP1 haploid cells
with pooled lentiviral sgRNA
library A or B
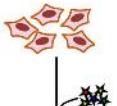
Phenotypic selection

CHIKV
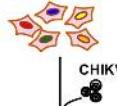
NGS of control and
selected cells

vs

**c**

Tanaka *et al.*
Loss-of-function genetic screen
in HAP1 human haploid cells

This study
CRISPR/Cas9 genetic screen
in HAP1 human haploid cells

192
(32.4%)

8
(1.4%)

192
(32.4%)

0
(0%)

0
(0%)

0
(0%)

200
(33.8%)

Zhang *et al.*
CRISPR/Cas9 genetic screen
in 3T3 mouse fibroblast

**b**
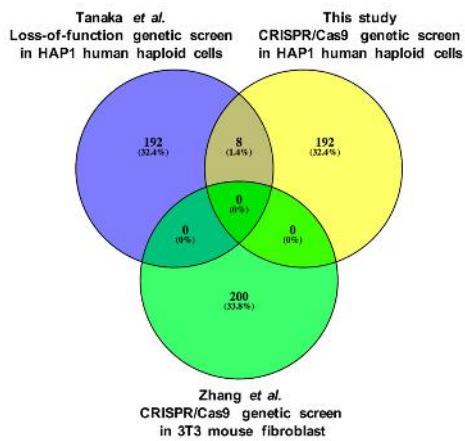
| Ranked Genes | MAGeCK rank | RIGER Rank |
|---|---|---|
| FHL1 | 1 | 1 |
| SLC35B2 | 2 | 2 |
| PAPSS1 | 3 | 3 |
| ELAVL1 | 4 | 4 |
| FURIN | 5 | 7 |
| YBX1 | 6 | 6 |
| GATAD1 | 7 | 5 |
| B3GAT3 | 8 | 9 |
| OR10W1 | 9 | 217 |
| MTRNR2L5 | 11 | 18 |
| COA5 | 12 | 436 |
| POMT2 | 13 | 641 |
| ELP5 | 14 | 47 |
| PRIM2 | 16 | 337 |
| NRCAM | 17 | 238 |
| PDE8B | 18 | 78 |
| EXT1 | 19 | 26 |
| ELF2 | 20 | 10 |
| MYT1L | 21 | 381 |
| CUL5 | 22 | 956 |
| MAP4K3 | 23 | 37 |
| ALDH18A1 | 24 | 107 |
| ACTR10 | 25 | 13544 |
| SLC7A6OS | 26 | 2704 |
| C11orf30 | 27 | 11 |
| BID | 28 | 8 |
| SMOC1 | 30 | 507 |
| SCGB1D1 | 32 | 84 |
| RPE65 | 33 | 75 |
| FAM124B | 34 | 459 |

**Extended Data Fig. 1 | CRISPR–Cas9 genetic screen identifies essential host factors of CHIKV infection. a**, Schematic of CRISPR–Cas9 genome-wide screen in HAP1 haploid cells. **b**, Ranked list of the top 30 genes identified using the MAGeCK algorithm and their corresponding rank in RIGER analysis. **c**, Venn diagram comparing the top 200 hits from our screen and previously published CRISPR and haploid screens[8,9] for CHIKV host factors.
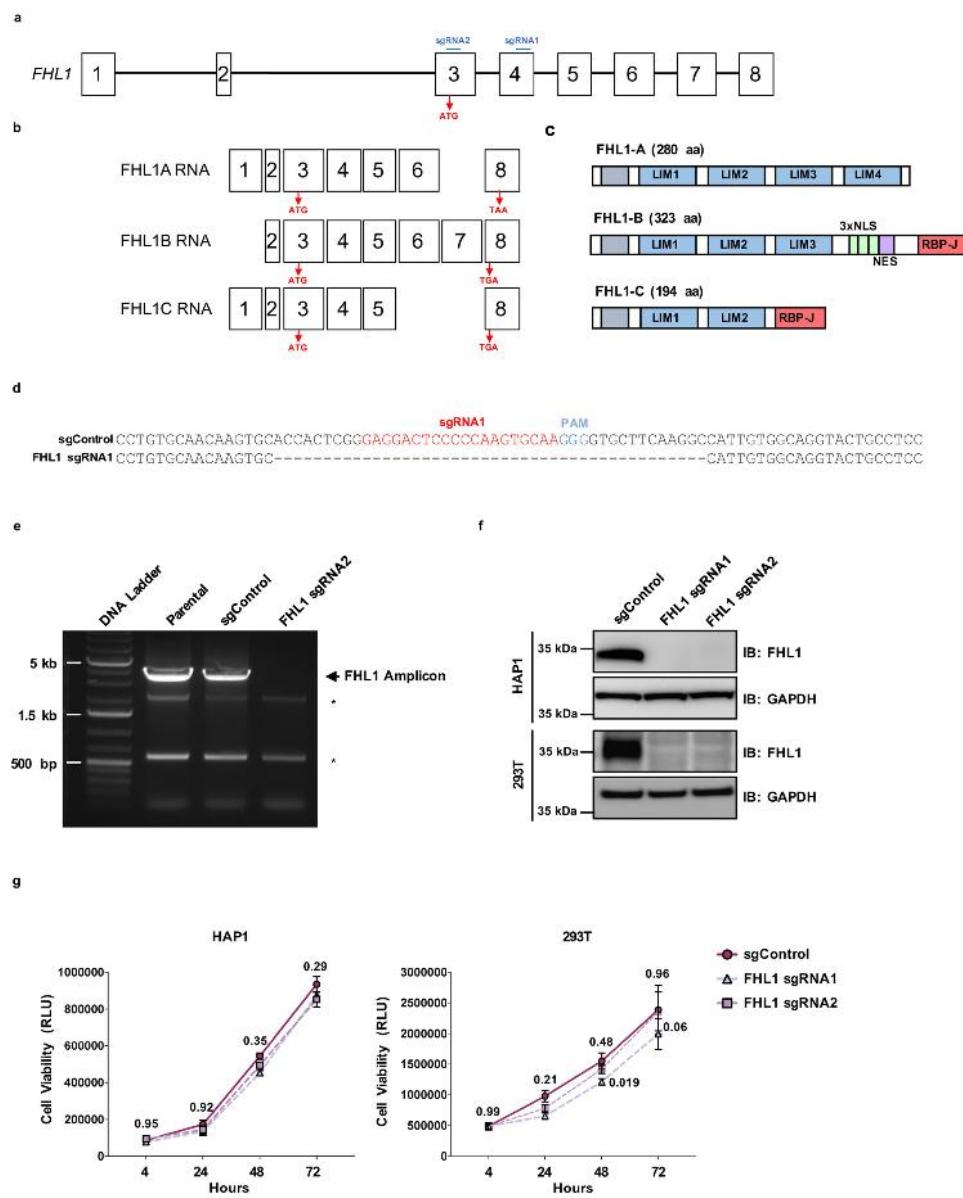
**Extended Data Fig. 2 | Validation of *FHL1* gene edition by CRISPR–Cas9. a–c**, Schematic of the genomic organization of *FHL1* (**a**), alternative splicing of the isoforms *FHL1A*, *FHL1B* and *FHL1C* (**b**) and their corresponding proteins (**c**). Initiation and stop codons are indicated in red and relative positions of the sequence targeted by the sgRNA are indicated in blue. **d**, Sanger sequencing of FHL1 in control and ΔFHL1 HAP1 cells. **e**, Genomic DNA was used for PCR amplification using primers flanking the sequence targeted by *FHL1* sgRNA2. The absence of an amplification product of 3.9 kb (black arrow) in the HAP1 clone suggests that a large indel is responsible for the absence of FHL1 expression. Asterisks indicate unspecific PCR products. Data are representative of two experiments. **f**, Immunoblot of FHL1 in control and ΔFHL1 cells. One representative experiment of three experiments is shown. **g**, Control and ΔFHL1 cells were plated and viability was assessed over a 72-h period using the CellTiter-Glo assay. Data shown are mean ± s.e.m. *n* = 2 independent experiments were performed in quadruplicate. Two-way ANOVA with Dunnett's multiple comparison test; *P* values are indicated.
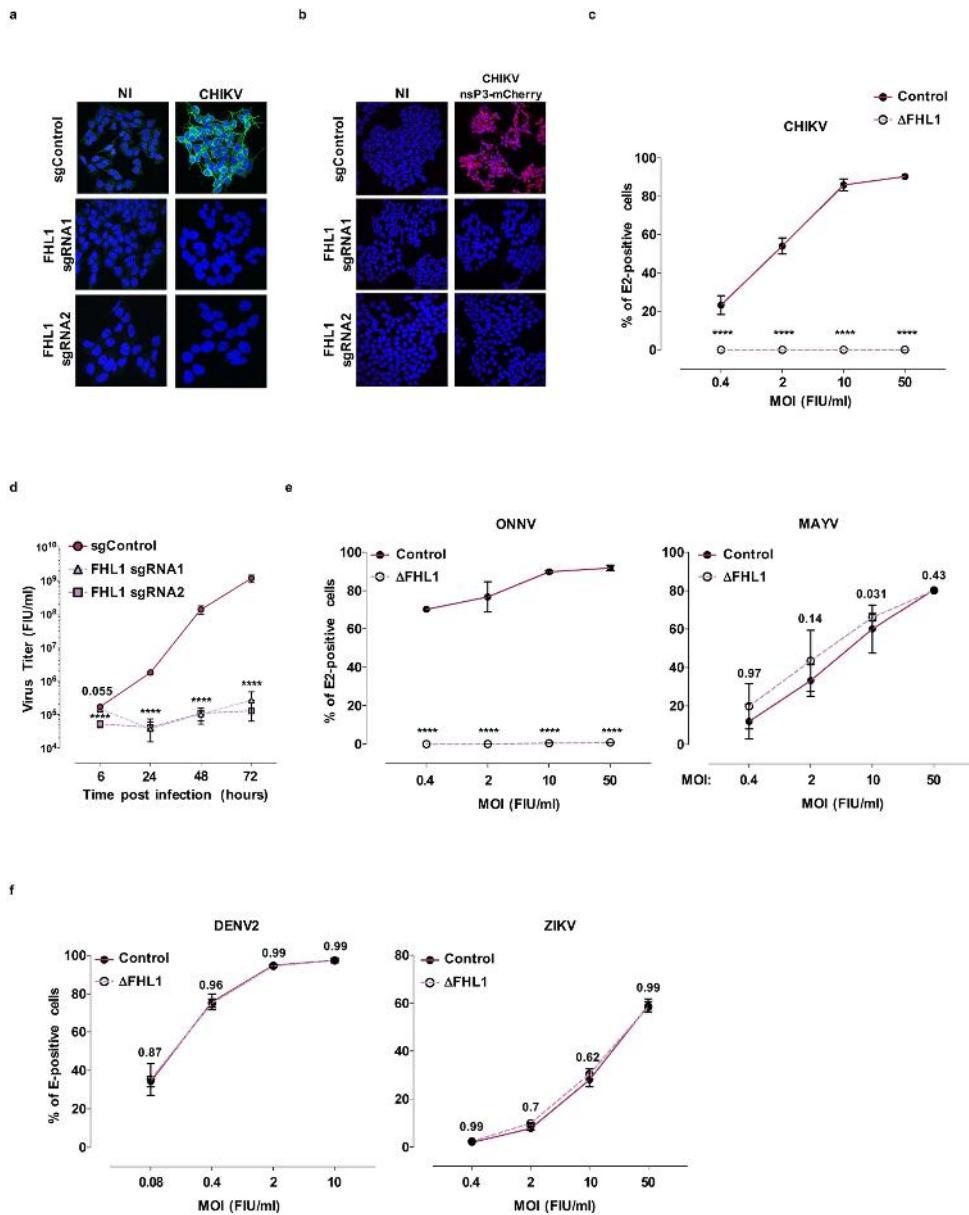
**Extended Data Fig. 3 | FHL1 is an essential host factor for CHIKV and ONNV infection. a**, Immunofluorescence images of control and ΔFHL1 HAP1 cells inoculated with CHIKV21 (MOI of 10), fixed 48 h after infection and stained for E2 expression. NI, not infected. **b**, Immunofluorescence images of control and ΔFHL1 HAP1 cells inoculated with CHIKV expressing nsP3–mCherry (MOI of 10) and fixed 48 h after infection. **a**, **b**, Images were taken on a fluorescence microscope and are representative of three experiments. **c**, Control and ΔFHL1 HAP1 cells were inoculated with increasing MOIs of CHIKV21, and infection was quantified 48 h after infection by flow cytometry using the anti-E2 3E4 monoclonal antibody. Data are mean ± s.d. $n = 3$ independent experiments performed in duplicate. Two-way ANOVA with Tukey's multiple comparison test. **d**, Multi-step growth curves with

the CHIKV21 strain in control or ΔFHL1 cells. Data are mean ± s.e.m. $n = 2$ independent experiments performed in duplicate. Two-tailed multiple $t$-tests with Holm–Sidak correction. **e**, Control and ΔFHL1 HAP1 cells were inoculated with increasing MOIs of ONNV or MAYV, and infection was quantified 48 h later by flow cytometry using anti-E2 3E4 and 265 monoclonal antibodies. Data are mean ± s.e.m. $n = 2$ independent experiments performed in duplicate. Two-way ANOVA with Tukey's multiple comparison test. **f**, Control and ΔFHL1 HAP1 cells were inoculated with increasing MOIs of DENV or ZIKV, and infection was quantified 48 h later by flow cytometry using the anti-E 4G2 monoclonal antibody. Data are mean ± s.e.m. $n = 3$ independent experiments performed in duplicate. Two-way ANOVA with Tukey's multiple comparison test. ****$P < 0.0001$.
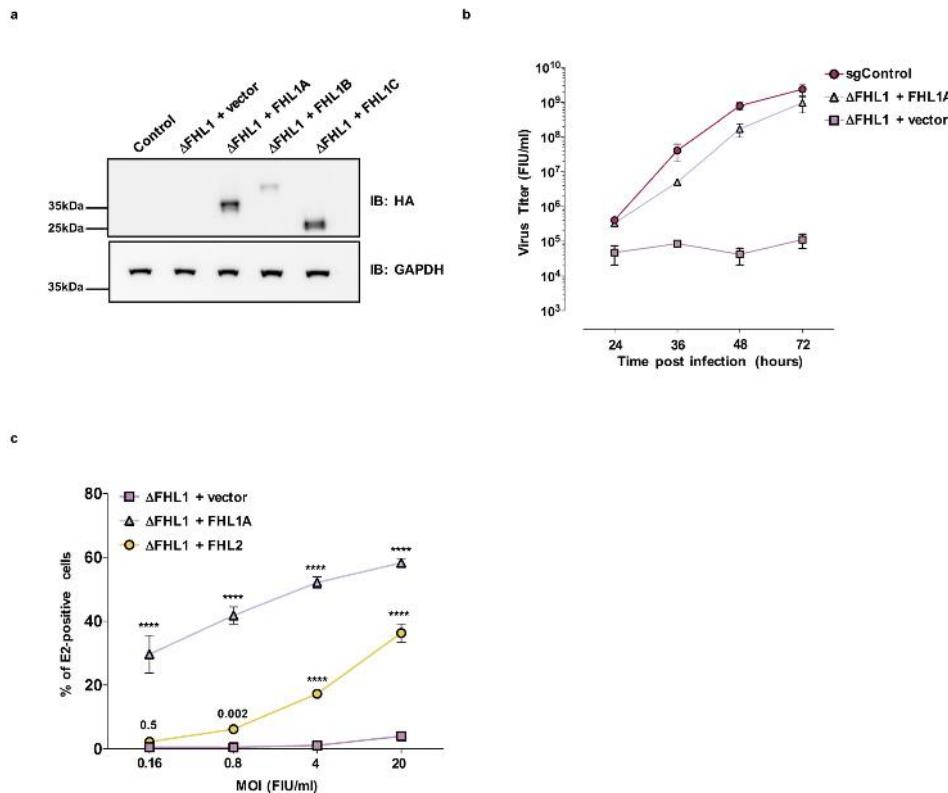
**Extended Data Fig. 4 | FHL1A and FHL2 expression in ΔFHL1 cells restores CHIKV infection. a,** Immunoblot (IB) of ectopic FHL1 expression in HAP1 cells stably transduced with an empty vector or FHL1A, FHL1B or FHL1C isoform. Data are representative of three experiments. **b,** Quantification in the supernatant of infected HAP1 cells of viral particles released by measuring the viral titre on Vero E6 cells. Data are representative of three experiments performed in duplicate. Data are mean ± s.e.m. **c,** ΔFHL1 HEK293T cells transfected with an empty vector or HA-tagged plasmids encoding FHL1A and FHL2 were infected with increasing MOIs of CHIKV21. Infection was quantified 24 h after infection by flow cytometry. Data are mean ± s.d. $n = 3$ experiments performed in duplicate. Two-way ANOVA with Dunnett's multiple comparison test. ****$P < 0.0001$.
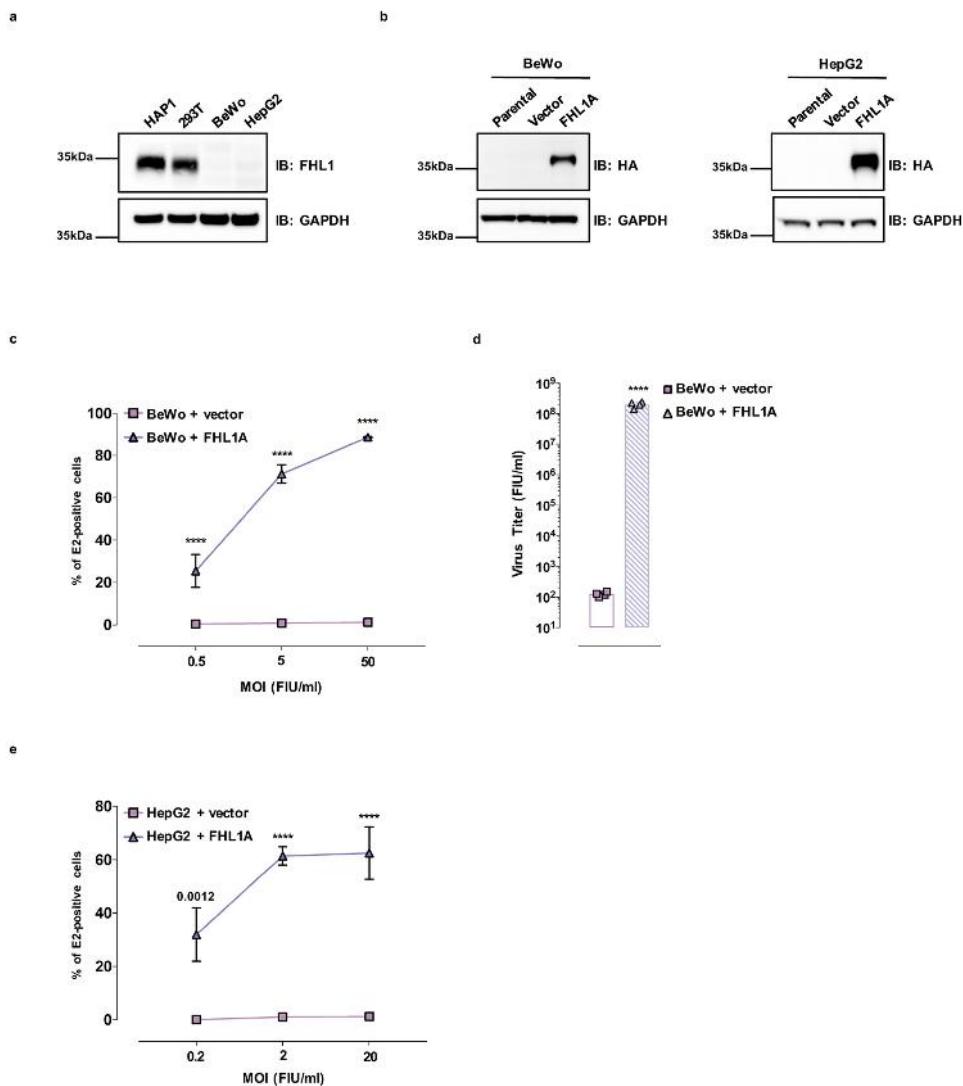
**Extended Data Fig. 5 | FHL1A overexpression in BeWo and HepG2 cells enhances CHIKV infection. a**, Expression of endogenous FHL1 in HAP1, HEK293T, BeWo and HepG2 cells. **b**, Immunoblot of ectopic FHL1 expression in Bewo and HepG2 cells stably transduced with an empty vector or HA-tagged FHL1A. **a, b**, Data are representative of three experiments. **c, d**, Bewo cells stably transduced with an empty vector or HA-tagged FHL1A were inoculated with increasing MOIs of CHIKV21. **c**, Infection was quantified 48 h after infection by flow cytometry using the anti-E2 3E4 monoclonal antibody. Data are mean ± s.e.m. $n = 3$ independent experiments performed in duplicate. Two-way ANOVA

with Tukey's multiple comparison test. **d**, Quantification of the viral particles released into the supernatants of infected cells, measured as the viral titre on Vero E6 cells. Data are mean ± s.d. $n = 2$ independent experiments performed in duplicate. Two-tailed Student's *t*-test. **e**, HepG2 cells stably transduced with an empty vector or FHL1A were inoculated with increasing MOIs of CHIKV-M-Gluc. Infection was quantified 48 h later as indicated in **c**. Data shown are mean ± s.e.m. $n = 2$ independent experiments performed in duplicate. Two-way ANOVA with Tukey's multiple comparison test. ****$P < 0.0001$.
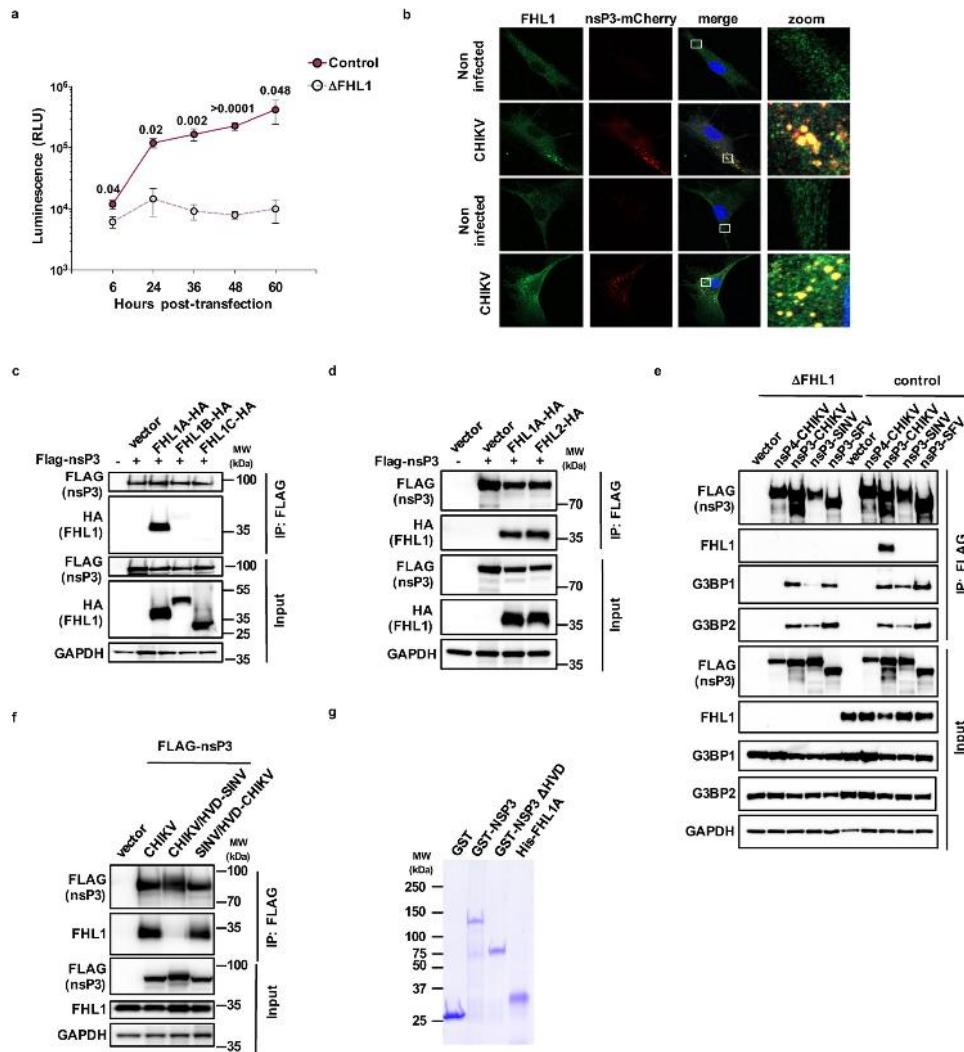
**Extended Data Fig. 6 | CHIKV nsP3 interacts with FHL1A and FHL2.**
**a**, Control or ΔFHL1 HAP1 cells were transfected with CHIKV-M-Gluc capped genomic RNA expressing *Gaussia* luciferase (Gluc). Gluc activity was monitored at the indicated time points. RLU, relative light units. Data are mean ± s.e.m. $n = 3$ independent experiments performed in quadruplicate. Two-tailed multiple *t*-tests with Holm-Sidak correction. **b**, Confocal microscopy of the colocalization of CHIKV nsP3 with FHL1 in fibroblasts inoculated with CHIKV nsP3–mCherry (MOI of 2), fixed 48 h after infection and stained with anti-FHL1 antibody. Images are representative of three experiments. **c**, Immunoassay of the interaction between CHIKV nsP3 and FHL1 isoforms in HEK293T cells transfected with Flag-tagged CHIKV nsP3 and either an empty vector or plasmids encoding the three HA-tagged FHL1 isoforms. Proteins from cell lysates were immunoprecipitated with anti-Flag antibody followed by immunoblot analysis with anti-Flag and anti-HA antibodies. **d**, Immunoassay of the interaction between CHIKV nsP3 and FHL2 in HEK293T cells transfected with Flag-tagged CHIKV nsP3 and either an

empty vector or plasmids encoding HA-tagged FHL1 and FHL2. Proteins from cell lysates were immunoprecipitated with anti-Flag followed by immunoblot analysis with anti-Flag and anti-HA. **e**, Endogenous FHL1, G3BP1 or G3BP2 immunoprecipitation from control and ΔFHL1 HEK293T cells transfected with plasmids encoding Flag-tagged CHIKV, Sindbis (SINV) or Semliki forest virus (SFV) nsP3. Proteins from cell lysates were immunoprecipitated with anti-Flag antibody followed by immunoblot analysis with anti-Flag, anti-FHL1, anti-G3BP1 and anti-G3BP2 antibodies. **f**, Endogenous FHL1 immunoprecipitation from HEK293T cells transfected with plasmids encoding Flag-tagged full length CHIKV nsP3, CHIKV nsP3 carrying the SINV HVD (CHIKV/HVD-SINV) or Sindbis nsP3 carrying CHIKV HVD (SINV/HVD-CHIKV). Proteins from cell lysates were immunoprecipitated with anti-Flag antibody followed by immunoblot analysis with anti-Flag and anti-FHL1 antibodies. **g**, Purified GST-tagged nsP3 constructs and HA-tagged FHL1A detected by Coomassie blue staining. **c–i**, One representative experiment of three experiments is shown.
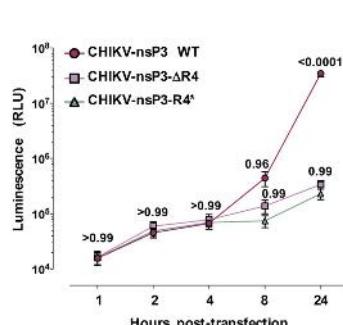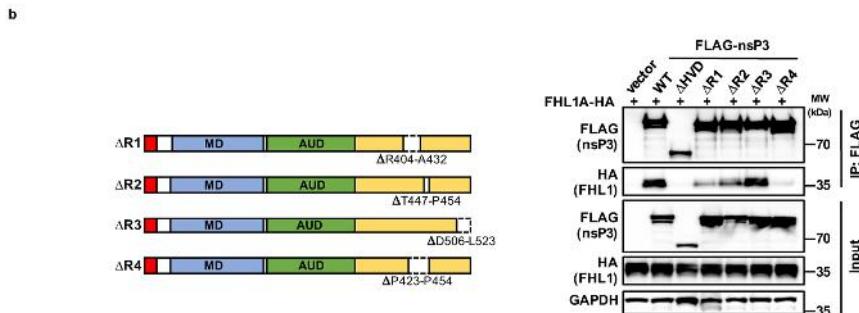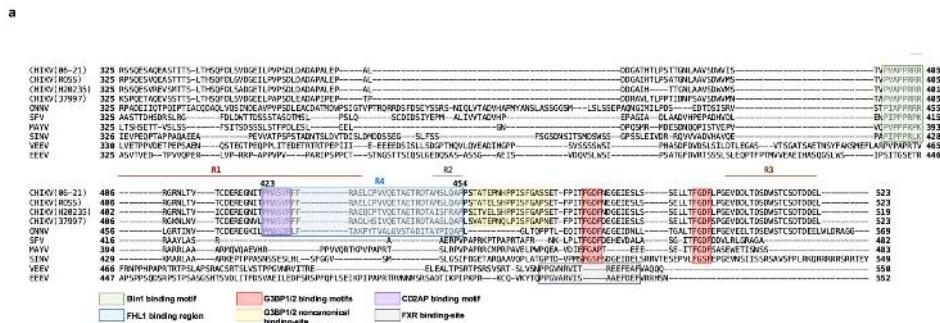
**Extended Data Fig. 7 | Mapping the FHL1–nsP3 interaction. a**, The sequence alignment of nsP3 protein HVD domains of representative members of New and Old World alphaviruses. Sequence alignment was performed with Clustal Omega and edited with Jalview. R1, R2 and R3 sequences of high homology between CHIKV strains and ONNV are defined by coloured lines. CHIKV(06–21) (GenBank accession number AM258992.1); CHIKV Ross (GenBank accession number MG280943.1); CHIKV H20235 (GenBank accession number MG208125.1); CHIKV 37997 (GenBank accession number AY726732.1); ONNV (GenBank accession number MF409176.1); SFV (GenBank accession number HQ848388.1); MAYV (GenBank accession number KY618137.1); SINV (GenBank accession number MF409178.1); EEEV (GenBank accession number Q4QXJ8.2); VEEV (GenBank accession number P27282.2). **b**, Left, schematic representation of CHIKV nsP3 constructs in which the R1, R2, R3 or R4 sequence was deleted. Right, HEK293T cells were transfected with FHL1A–HA and either an empty vector or plasmids encoding Flag-tagged nsP3 constructs. Cell lysates were immunoprecipitated with anti-Flag antibody followed by immunoblot analysis with anti-HA or anti-Flag antibodies. One experiment representative of three experiments is shown. **c**, Alignment of nsP3 regions containing the wild-type R4 sequence or the corresponding randomized sequence. Dashes represent identical amino acids. **d**, Control HEK293T cells were transfected with the indicated CHIKV capped in vitro transcribed RNA-expressing *Renilla* luciferase (Rluc). Rluc activity was monitored at indicated time points. RLU, relative light units. Data are mean $\pm$ s.e.m. $n = 2$ independent experiments performed in quadruplicate. Two-way ANOVA with Tukey's multiple comparison test.
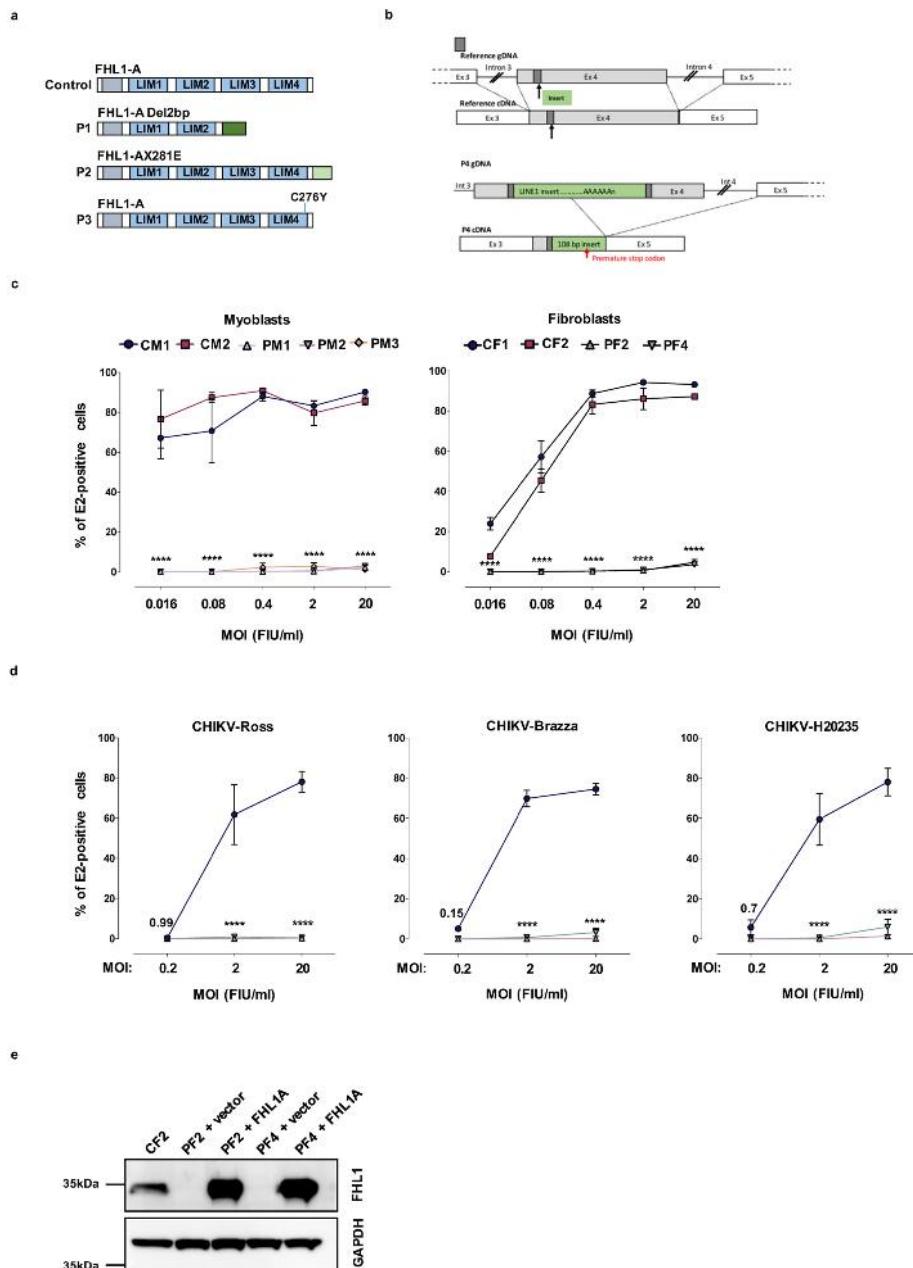
**Extended Data Fig. 8 | CHIKV infection of myoblasts and fibroblasts derived from patients with EDMD. a**, Schematic of FHL1A proteins from three patients with EDMD (P1, P2 and P3). **b**, Schematic of FHL1 genomic organization in the patient with a LINE1 insertion within exon 4 (P4). **c**, Myoblasts and fibroblasts from patients with EDMD or healthy donors were infected with increasing MOIs of CHIKV21, and infection was quantified 24 h later by flow cytometry using the anti-E2 3E4 monoclonal antibody. Data are mean ± s.e.m. $n = 2$ experiments performed in duplicate for myoblasts; $n = 3$ independent experiments performed in duplicate for fibroblasts. Two-way ANOVA with Dunnett's multiple comparison test. **d**, Fibroblasts from patients with EDMD or

healthy donors were inoculated with increasing MOIs of CHIKV Ross, CHIKV Brazza, CHIKV H20235, and infection was quantified 24 h later by flow cytometry using the anti-E2 3E4 monoclonal antibody. Data are mean ± s.e.m. $n = 3$ independent experiments performed in duplicate. Two-way ANOVA with Dunnett's multiple comparison test. **e**, Immunoblot of ectopic FHL1 expression in primary fibroblasts (PF2 and PF4) obtained from patients that were stably transduced with an empty vector or a plasmid encoding HA–FHL1A. One representative of two experiments performed in duplicate is shown. Data are mean ± s.d. ****$P < 0.0001$.

**Extended Data Fig. 9 | Mouse FHL1 interacts with CHIKV nsP3 and restores infection in ΔFHL1 cells. a**, Sequence alignment of mouse and human FHL1A proteins. **b**, HEK293T cells were co-transfected with Flag-tagged CHIKV nsP3 and plasmids encoding HA-tagged mouse FHL1 (mFHL1) or human FHL1A (hFHL1A). Proteins from cell lysates were immunoprecipitated with anti-HA antibody followed by immunoblot analysis with anti-Flag (nsP3) and anti-HA (FHL1) antibodies. **c**, Immunoblot of FHL1 ectopic expression in ΔFHL1 HEK293T cells stably transduced with a plasmid encoding mouse FHL1 or human FHL1A. **b**, **c**, Data are representative of three experiments. **d**, Cells showed in

**c** were inoculated with increasing MOIs of CHIKV21. Infection was quantified by flow cytometry at 24 h after infection using anti-E2 3E4 monoclonal antibody. Data are mean ± s.d. $n = 3$ independent experiments performed in duplicate. Two-way ANOVA with Dunnett's multiple comparison test. **e**, Left, immunoblot of endogenous FHL1 in control and ΔFHL1 C2C12 mouse cells. Right, control and ΔFHL1 cells were inoculated with CHIKV21 or MAYV (MOI of 2) and infection was quantified at 24 h after infection by flow cytometry using anti-E2 3E4 or anti-E2 265 monoclonal antibodies. One representative experiment of three experiments is shown. ****$P < 0.0001$.

Corresponding author(s):    Ali Amara

Last updated by author(s):    Jun 14, 2019

# Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist .

## Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

| n/a | Confirmed | |
|---|---|---|
| ☐ | ☒ | The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement |
| ☐ | ☒ | A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly |
| ☐ | ☒ | The statistical test(s) used AND whether they are one- or two-sided<br>*Only common tests should be described solely by name; describe more complex techniques in the Methods section.* |
| ☐ | ☒ | A description of all covariates tested |
| ☐ | ☒ | A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons |
| ☐ | ☒ | A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals) |
| ☒ | ☐ | For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted<br>*Give P values as exact values whenever suitable.* |
| ☒ | ☐ | For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings |
| ☒ | ☐ | For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes |
| ☒ | ☐ | Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated |

*Our web collection on statistics for biologists contains articles on many of the points above.*

## Software and code

Policy information about availability of computer code

| Data collection | Invitrogen Attune NxT software |
|---|---|
| Data analysis | Graph Prism 7.0<br>Flowjow 10.5.2<br>MAGeCK 0.5.4<br>RIGER, as implemented in GENE-E version 3.0.215(https://software.broadinstitute.org/GENE-E/) |

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

## Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

All data will be available on request from the authors.

# Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

# Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

| | |
|---|---|
| Sample size | No experiments presented in this study required sample size to be determined |
| Data exclusions | No data were excluded |
| Replication | All cell culture experiments were repeated at least two independent times. For these experiments cells were inoculated with different multiplicity of infection (MOI) and data presented in figures showed results for one representative MOI in the linear range of infection. All attempts at replication were successful |
| Randomization | No experiments presented in this study required randomization |
| Blinding | No experiments presented in this study required blinding |

# Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

### Materials & experimental systems

| n/a | Involved in the study |
|---|---|
| ☐ | ☒ Antibodies |
| ☐ | ☒ Eukaryotic cell lines |
| ☒ | ☐ Palaeontology |
| ☐ | ☒ Animals and other organisms |
| ☐ | ☒ Human research participants |
| ☒ | ☐ Clinical data |

### Methods

| n/a | Involved in the study |
|---|---|
| ☒ | ☐ ChIP-seq |
| ☐ | ☒ Flow cytometry |
| ☒ | ☐ MRI-based neuroimaging |

## Antibodies

| | |
|---|---|
| Antibodies used | anti-FHL1 mAb (ref MAB5938, Biotchene), polyclonal anti-FHL1 rabbit Ab (ref NBP1-88745, Novus Biologicals) anti-GAPDH mAb (ref SC-47724, Santa Cruz Biotechnology), polyclonal rabbit anti-HA (ref 3724, Cell Signaling Technology), anti-FLAG M2 mAb (ref F1804, SIGMA), anti-RFP (ref 6G6, Chromotek), anti-CHIKV E2 mAb (3E4 was provided  by V. Choumet, Insitut Pasteur), anti-alphavirus E2 mAb (CHIK-265 was provided by Michael Diamonds, University school of medicine, St Louis, USA), anti-vimentin antibody (ab24525, abcam), anti-EEEV E1 mAb (ref MAB8754, Sigma), anti-pan-flavivirus E protein mAb (4G2), anti-dsRNA J2 mAb (Scicons), Alexa FluorTM 488-conjugated goat anti-rabbit IgG (A11034, Invitrogen), Alexa FluorTM-647-conjugated goat anti-chicken IgG (ab150175, abcam),  Alexa FluorTM 488-conjugated goat anti-mouse IgG (115-545-003, Jackson ImmunoResearch),  Alexa FluorTM 647-conjugated goat anti-mouse IgG (115-606-062, Jackson ImmunoResearch), peroxydase-conjugated donkey anti-rabbit IgG (711-035-152, Jackson ImmunoResearch), and anti-mouse/HRP (P0260, Dako Cytomotion) |
| Validation | Anti-FHL1 antibody was validated by Western Blot and immofluoresence assay. All the antibody used in this study have been already validated by manufacturers. |

## Eukaryotic cell lines

Policy information about cell lines

| | |
|---|---|
| Cell line source(s) | HAP1 cells were purchased from Horizon Discovery. 293FT were purchased from Thermo Fisher Scientific (R70007). HEK-293T and VeroE6 were purchased from ATCC. HepG2 and BHK21 were provided by Olivier Schwartz, (Institut Pasteur, France). Primary myoblasts and primary fibroblasts were provided by G. Bonne, R. Ben Yaou and A. Bertrand Legout (Institut de Myologie, France). Human placenta choriocarcinoma Bewo cells were provided by M. Lecuit (Institut Pasteur, France). AP61 mosquito (Aedes pseudoscutellaris) cells were provided by P. Despres (Institut Pasteur, France) |
| Authentication | none |
| Mycoplasma contamination | All cell lines were tested negative for mycoplasma contamination using the MycoAlert Myoplasma detection kit from Lonza Primary cell lines were not tested |
| Commonly misidentified lines (See ICLAC register) | No misidentified cell lines were used in this study |

# Animals and other organisms

Policy information about studies involving animals; ARRIVE guidelines

# LETTER

# The fungal mycobiome promotes pancreatic oncogenesis via activation of MBL

Berk Aykut[1,7], Smruti Pushalkar[2,7], Ruonan Chen[1], Qianhao Li[2], Raquel Abengozar[1], Jacqueline I. Kim[1], Sorin A. Shadaloey[1], Dongling Wu[1], Pamela Preiss[1], Narendra Verma[3], Yuqi Guo[2], Anjana Saxena[4,5], Mridula Vardhan[2], Brian Diskin[1], Wei Wang[1], Joshua Leinwand[1], Emma Kurz[1], Juan A. Kochen Rossi[1], Mautin Hundeyin[1], Constantinos Zambrinis[1], Xin Li[2], Deepak Saxena[1,2,8]* & George Miller[1,6,8]*

Bacterial dysbiosis accompanies carcinogenesis in malignancies such as colon and liver cancer, and has recently been implicated in the pathogenesis of pancreatic ductal adenocarcinoma (PDA)[1]. However, the mycobiome has not been clearly implicated in tumorigenesis. Here we show that fungi migrate from the gut lumen to the pancreas, and that this is implicated in the pathogenesis of PDA. PDA tumours in humans and mouse models of this cancer displayed an increase in fungi of about 3,000-fold compared to normal pancreatic tissue. The composition of the mycobiome of PDA tumours was distinct from that of the gut or normal pancreas on the basis of alpha- and beta-diversity indices. Specifically, the fungal community that infiltrated PDA tumours was markedly enriched for *Malassezia* spp. in both mice and humans. Ablation of the mycobiome was protective against tumour growth in slowly progressive and invasive models of PDA, and repopulation with a *Malassezia* species—but not species in the genera *Candida*, *Saccharomyces* or *Aspergillus*—accelerated oncogenesis. We also discovered that ligation of mannose-binding lectin (MBL), which binds to glycans of the fungal wall to activate the complement cascade, was required for oncogenic progression, whereas deletion of MBL or C3 in the extratumoral compartment—or knockdown of C3aR in tumour cells—were both protective against tumour growth. In addition, reprogramming of the mycobiome did not alter the progression of PDA in *Mbl*- (also known as *Mbl2*) or *C3*-deficient mice. Collectively, our work shows that pathogenic fungi promote PDA by driving the complement cascade through the activation of MBL.

It has recently been reported that intrapancreatic bacteria expand by about 1,000-fold in PDA[1]. Here we show that there is a similar and marked increase in intratumoral fungi in PDA and in mouse models of this disease (Fig. 1a–d). Because there is direct communication between the gut and pancreatic duct via the sphincter of Oddi, we postulated that endoluminal fungi can access the pancreas. To test this, we administered *Saccharomyces cerevisiae* labelled with green fluorescent protein (GFP) to control and tumour-bearing mice via oral gavage. Fungi migrated into the pancreas within 30 min, which suggests that the gut mycobiome can directly influence the pancreatic microenvironment (Fig. 1e).

We next assessed whether there is evidence of fungal dysbiosis during tumorigenesis, using *p48^{cre};LSL-Kras^{G12D}* (*p48* is also known as *Ptf1a*) mice (hereafter referred to as KC mice), which express oncogenic *Kras* in their pancreatic progenitor cells and are a model for the development of slowly progressive PDA[2]. A comparison between the fungal communities of the gut and within the pancreas in 30-week-old KC mice, by principal coordinate analysis (PCoA), suggested that the mycobiomes of the gut and tumours clustered separately (Fig. 1f). We also observed reduced alpha-diversity in the transformed pancreas compared with

the gut (Fig. 1g). Ascomycota and Basidiomycota were the only phyla that were detected in pancreatic tissue, whereas Mortierellomycota and Mucoromycota were also detected in the gut at a low abundance (Fig. 1h). The most-prevalent genus in the pancreata of KC mice was *Malassezia*, at about 20% abundance; this represents a marked increase in relative abundance compared to the presence of this genus in the gut (Fig. 1i). Of note, benign pancreatic inflammation did not increase fungal infiltration into the pancreas (Extended Data Fig. 1).

To determine whether the gut mycobiome is reprogrammed during the course of oncogenesis, we performed a longitudinal analysis of faecal samples from KC mice and littermate controls. PCoA suggested that, whereas wild-type and KC mice had similar fungal communities early in life, by 30 weeks of age there were differences in beta-diversity between the gut mycobiomes in the two backgrounds (Fig. 1j–l). Accordingly, fungal communities in the gut of KC and wild-type mice differed considerably at 30 weeks (Extended Data Fig. 2).

We next analysed the faecal and tumour mycobiome in patients with PDA. As in mice, Ascomycota and Basidiomycota were the most common phyla in the gut and in tumour tissue of humans (Fig. 2a). At the genus level (and once again parallel to our mice data), *Malassezia* was more prevalent in tumour tissues than in the gut (Fig. 2b). Moreover, alpha-diversity analyses revealed considerable differences between the gut and PDA-tumour tissue in humans (Fig. 2c). PCoA confirmed that there were distinct clusters of fungal communities in the tumour tissue and gut of patients with PDA (Fig. 2d). Furthermore, the mycobiome in pancreata from patients with PDA clustered separately from that in the pancreata of healthy individuals (Fig. 2e). Collectively, these data indicate that the mycobiome of PDA tumours is distinct from that of the gut or healthy pancreas.

To determine the influence of fungal dysbiosis on the progression of PDA, we ablated the mycobiome using oral administration of amphotericin B in the KC mouse model. Ablation of the mycobiome protected the mice against oncogenic progression (Fig. 3a). Similarly, amphotericin B was protective against tumour progression in an aggressive orthotopic model of PDA that uses tumour cells derived from *Pdx1^{cre};Kras^{G12D};Tp53^{R172H}* (*Tp53* is also known as *Trp53*) mice (hereafter, KPC mice)[3] (Fig. 3b). Ablation of the mycobiome potentiated the effect of chemotherapy based on gemcitabine (Fig. 3c). Of note, treatment with fluconazole was also protective against tumour progression (Extended Data Fig. 3a). However, treatment with antifungal agents did not offer protection against tumour growth in germ-free mice (Extended Data Fig. 3b). Furthermore, consistent with absence of increased fungal infiltration in pancreatitis, treatment with antifungal agents did not ameliorate benign pancreatic inflammation (Extended Data Fig. 3c–e).

To confirm that fungal dysbiosis accelerates the progression of PDA, we repopulated mice treated with amphotericin B with *Malassezia*

[1]S. Arthur Localio Laboratory, Department of Surgery, New York University School of Medicine, New York, NY, USA. [2]Department of Basic Science and Craniofacial Biology, New York University College of Dentistry, New York, NY, USA. [3]Department of Medicine, New York University School of Medicine, New York, NY, USA. [4]Biology Department, Brooklyn College, CUNY, New York, NY, USA. [5]Biology and Biochemistry Programs, Graduate Center CUNY, New York, NY, USA. [6]Department of Cell Biology, New York University School of Medicine, New York, NY, USA. [7]These authors contributed equally: Berk Aykut, Smruti Pushalkar. [8]These authors jointly supervised this work: Deepak Saxena, George Miller. *e-mail: ds100@nyu.edu; george.miller@nyumc.org
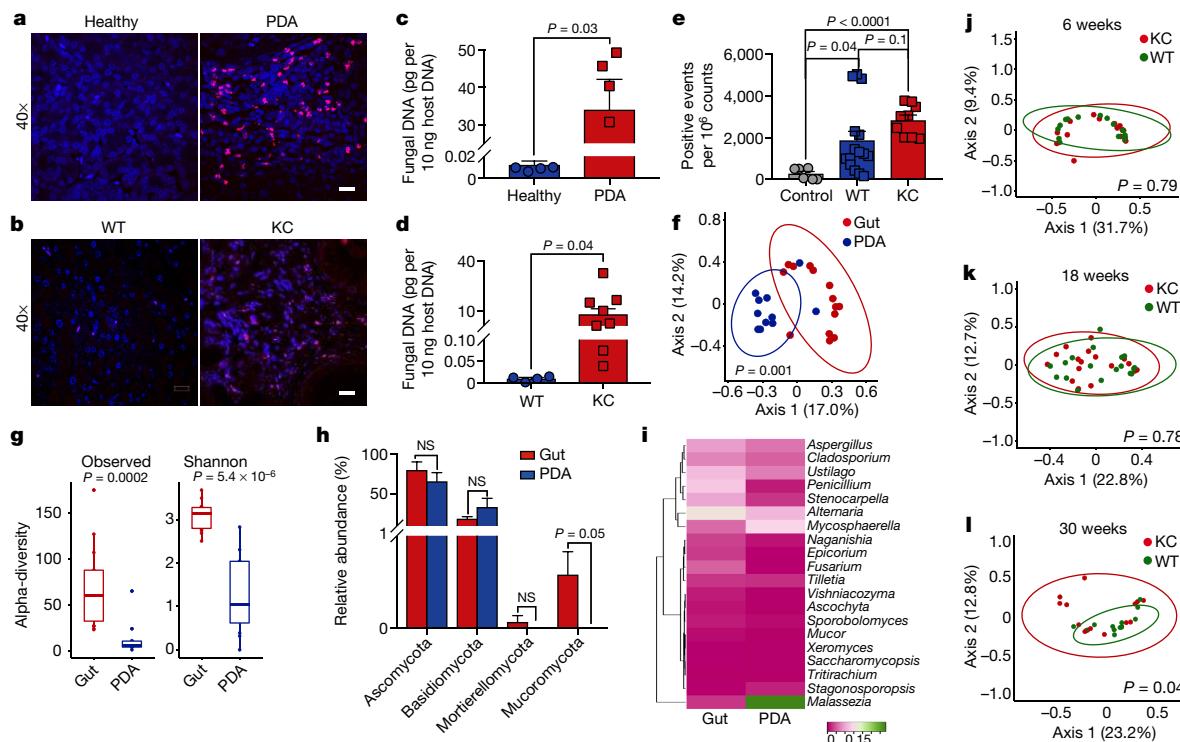
**Fig. 1 | PDA is characterized by a distinctive intratumoral and gut mycobiome. a**, The abundance of intrapancreatic fungi was compared between healthy individuals and patients with PDA who were matched for age, gender and body mass index, using fluorescent in situ hybridization (FISH). $n = 3$ individuals per group. Representative images are shown. Scale bar, 20 μm. **b**, The abundance of intrapancreatic fungi was compared in three-month-old, littermate wild-type (WT) and KC mice by FISH. Representative images are shown. $n = 3$ mice per group. Scale bar, 20 μm. **c**, Fungal DNA content was compared in the pancreata of healthy individuals and patients with PDA who were matched for age, gender and body mass index, using quantitative PCR (qPCR). **d**, Fungal DNA content was compared in the pancreata of three-month-old wild-type and KC mice, using qPCR. **e**, GFP-labelled *S. cerevisiae* was administered to wild-type ($n = 15$) and KC ($n = 9$) mice via oral gavage. Pancreata were collected at 30 min, and the number of GFP+ foci was determined by flow cytometry in comparison to mock-treated mice (control, $n = 6$ mice). This experiment was repeated twice. **f–i**, The guts and intrapancreatic ($n = 14$ and 11 biologically independent samples, respectively) mycobiomes of

30-week-old KC mice were analysed by 18S internal transcribed space (ITS) sequencing. **f**, PCoA plots based on a Bray–Curtis dissimilarity matrix. Each symbol represents a sample from the gut (red) or pancreas (blue). Clusters were determined by pairwise permutational analysis of variance (PERMANOVA). The *x* and *y* axes indicate variation (%), and the ellipses indicate the 95% confidence interval. **g**, The gut and intrapancreatic mycobiomes in 30-week-old KC mice were analysed for alpha-diversity measures, including observed operational taxonomic units (OTUs) and Shannon indices. Box plots show median, 25th and 75th percentiles, and whiskers that extend to 1.5× the interquartile range. **h**, Taxonomic composition of mycobiota assigned to the phylum level, on the basis of their average relative abundance. NS, not significant. **i**, Heat map showing log$_2$-transformed relative abundancies of the top 20 fungal genera in the gut and pancreas. **j–l**, PCoA plots of fungal communities in faeces of 6- (**j**), 18- (**k**) and 30- (**l**) week-old wild-type and KC mice, based on a Bray–Curtis dissimilarity matrix, as in **f**. Data in **c–e**, **h** are mean ± s.e.m. *P* values determined by two-tailed Student's *t*-test (**c–e**, **h**), pairwise PERMANOVA (**f**, **j–l**) or two-sided Wilcoxon rank-sum test (**g**).

*globosa*, which is present at an increased abundance in PDA and in mouse models of this cancer (Figs. 1i, 2b). Of note, the *M. globosa* ATCC strain that we used in our repopulation experiments had 100% sequence identity to the *Malassezia* taxon that was the most abundant in PDA (Supplementary Table 1). Control mice were repopulated with *Candida* sp., *S. cerevisiae* or *Aspergillus* sp. or treated with vehicle. Of these, only *M. globosa* accelerated the growth of PDA tumours; the other taxa, and treatment with vehicle, had no effect (Fig. 3d). Repopulation with *Candida tropicalis* also did not accelerate the growth of PDA tumours (Extended Data Fig. 3f).

MBL is a mannose-binding lectin that recognizes fungal pathogens and activates the lectin pathway of the complement cascade[4]. Expression of *MBL* (also known as *MBL2*) was associated with reduced survival in patients with PDA, on the basis of transcriptomic data from The Cancer Genome Atlas (TCGA) (Extended Data Fig. 4a). We postulated that fungi may promote tumorigenesis via activation of MBL. Accordingly, MBL-null KC mice exhibited delayed oncogenic progression (Fig. 4a). Deletion of *Mbl* was also protective against the growth of orthotopic tumours of KPC cells, and resulted in extended survival of the mice (Fig. 4b, c). Moreover, treatment with amphotericin B did not provide protection against tumour growth in MBL-null mice (Extended Data Fig. 4b). Similarly, *Malassezia*—which binds C-type

lectin receptors[5]—did not accelerate tumour progression in MBL-null mice (Extended Data Fig. 4c).

The C3 complement cascade has previously been investigated in PDA and other cancers, and is potently oncogenic by diverse mechanisms that include increasing the proliferation motility and invasiveness of tumour cells, and corrupting adaptive immune responses[6,7]. Because MBL initiates the lectin pathway of the complement cascade that triggers C3 convertase, we postulated that the fungus–MBL axis promotes the progression of PDA via complement activation. Similar to *MBL*, the expression of *C3* was associated with a trend towards reduced survival in patients with PDA (Extended Data Fig. 4d). We found robust expression of C3a in the pancreata of KC mice, and this was nearly absent in wild-type or MBL-null KC mice (Extended Data Fig. 4e). Consistent with our hypothesis, recombinant C3a accelerated the proliferation of KPC cells in vitro (Extended Data Fig. 4f) and the growth of KPC tumours in vivo (Fig. 4d), whereas *C3*-deficient mice were protected against PDA progression (Fig. 4e). Similarly, knockdown of C3aR in PDA cells (Fig. 4f) mitigated tumour growth (Fig. 4g). Moreover, we found that targeting the mycobiome had no additional effect in C3-deficient animals (Fig. 4h). In aggregate, these data indicate that the pancreatic mycobiome requires the MBL–C3 axis to promote tumour growth.
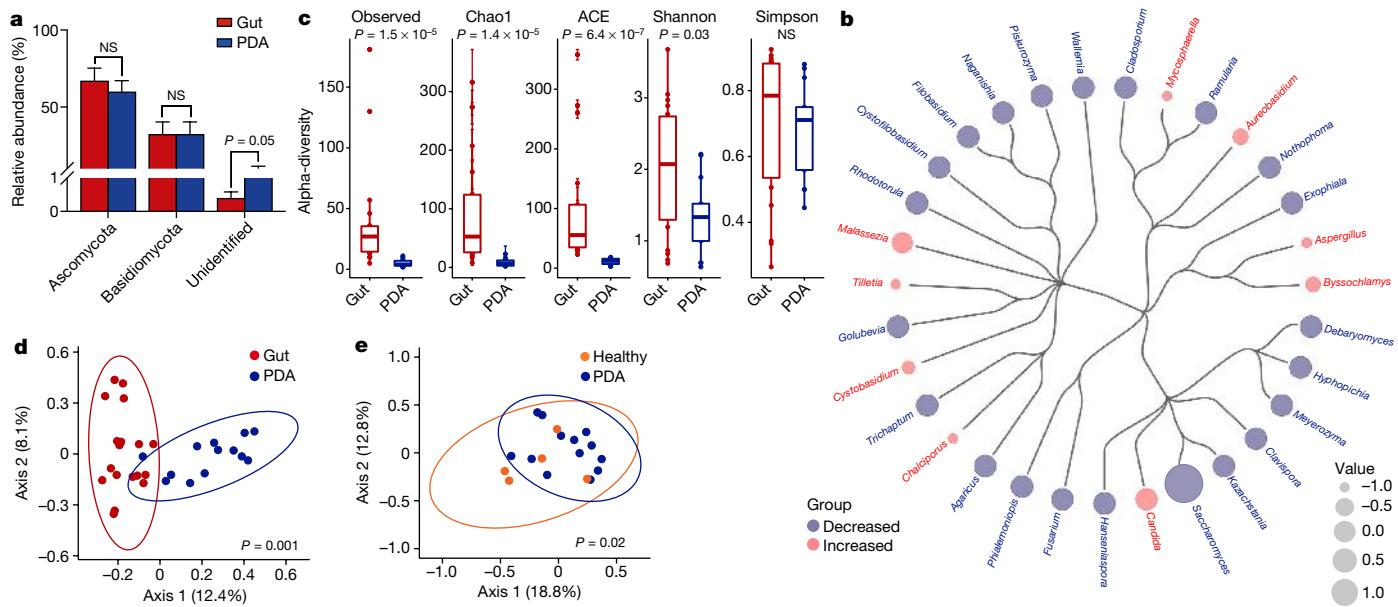
**Fig. 2 | PDA in humans is associated with a distinct mycobiome.**
**a**–**d**, Gut and tumour (n = 18 and 13 biologically independent specimens, respectively) specimens from patients with PDA were analysed by 18S ITS sequencing. **a**, Taxonomic composition of mycobiota assigned to the phylum level, on the basis of their average relative abundance. Data are mean ± s.e.m. **b**, Hierarchical tree cladogram, depicting differences between the gut and tumours in terms of the taxonomic composition of mycobiota assigned to the genus level (on the basis of their average relative abundance). **c**, The gut and tumour mycobiomes of patients with

PDA were tested for alpha-diversity measures, including observed OTUs, abundance-based coverage estimates (ACE), and the Chao1, Shannon and Simpson indices. Box plots as in Fig. 1g. **d**, PCoA plots of gut (n = 18) and intratumoral (n = 13) fungal communities in patients with PDA, based on a Bray–Curtis dissimilarity matrix, as in Fig. 1f. **e**, PCoA plots of fungal communities in pancreata of patients with PDA (n = 13) and healthy individuals (n = 5), based on a Bray–Curtis dissimilarity matrix. P values determined by two-tailed Student's t-test (**a**), two-sided Wilcoxon rank-sum test (**c**) or pairwise PERMANOVA (**d**, **e**).
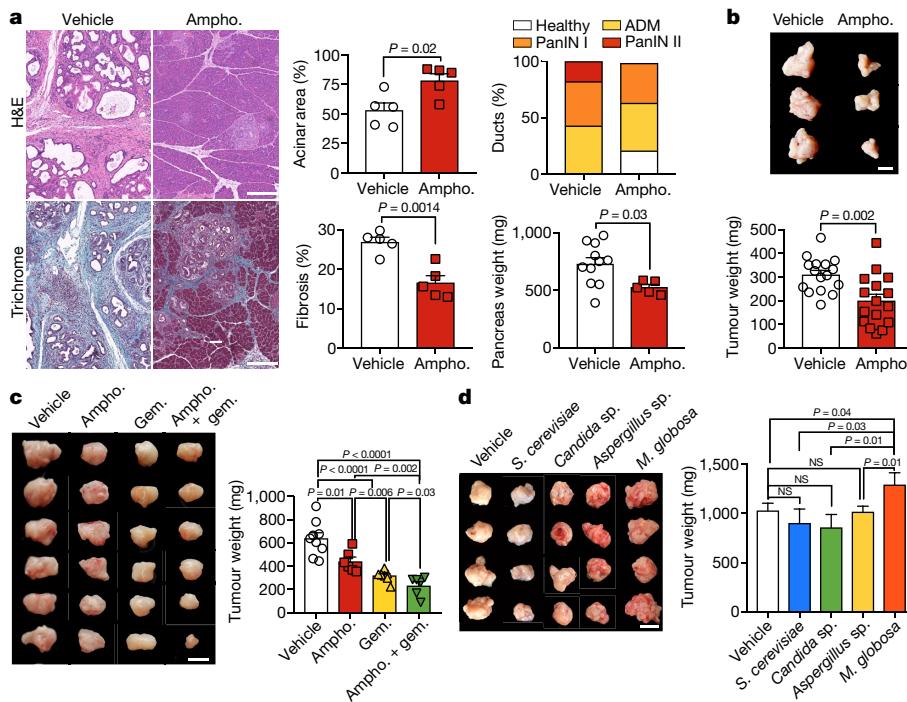


**Fig. 3 | Fungal dysbiosis promotes pancreatic oncogenesis. a**, KC mice treated with amphotericin B (ampho.) or vehicle were killed at three months old. Pancreatic weights (n = 5 and 11 mice treated with amphotericin B and vehicle, respectively) were recorded. Representative sections stained with haematoxylin and eosin (H&E) or trichrome. The percentage of preserved acinar area, and the fraction of normal ducts, acinoductal metaplasia (ADM) and graded (I and II) pancreatic intra-epithelial neoplasia (PanIN) lesions were determined on the basis of H&E staining. The fraction of fibrotic area per pancreas was calculated on the basis of trichrome staining. **b**, Wild-type mice that bear orthotopic PDA tumours were treated with vehicle or amphotericin B (n = 16 mice per group, data pooled from 3 independent experiments) and killed three

weeks later. Tumours were collected and weighed. Data are representative of at least five experiments. **c**, Wild-type mice that bear orthotopic PDA tumours were treated with vehicle (n = 9 mice), amphotericin B (n = 6 mice), gemcitabine (gem.) (n = 8 mice) or amphotericin B and gemcitabine (n = 6 mice). Tumour weight was recorded after three weeks of treatment. **d**, Wild-type mice treated with amphotericin B were repopulated with M. globosa (n = 8 mice), S. cerevisiae (n = 9 mice), Candida sp. (n = 8 mice), Aspergillus sp. (n = 10 mice) or vehicle (n = 8 mice), and killed three weeks later. Tumours were collected and weighed. Data are representative of two experiments. Scale bars, 200 μm (**a**), 1 cm (**b**–**d**). Data are mean ± s.e.m. P values determined by two-tailed Student's t-test (**a**–**d**).
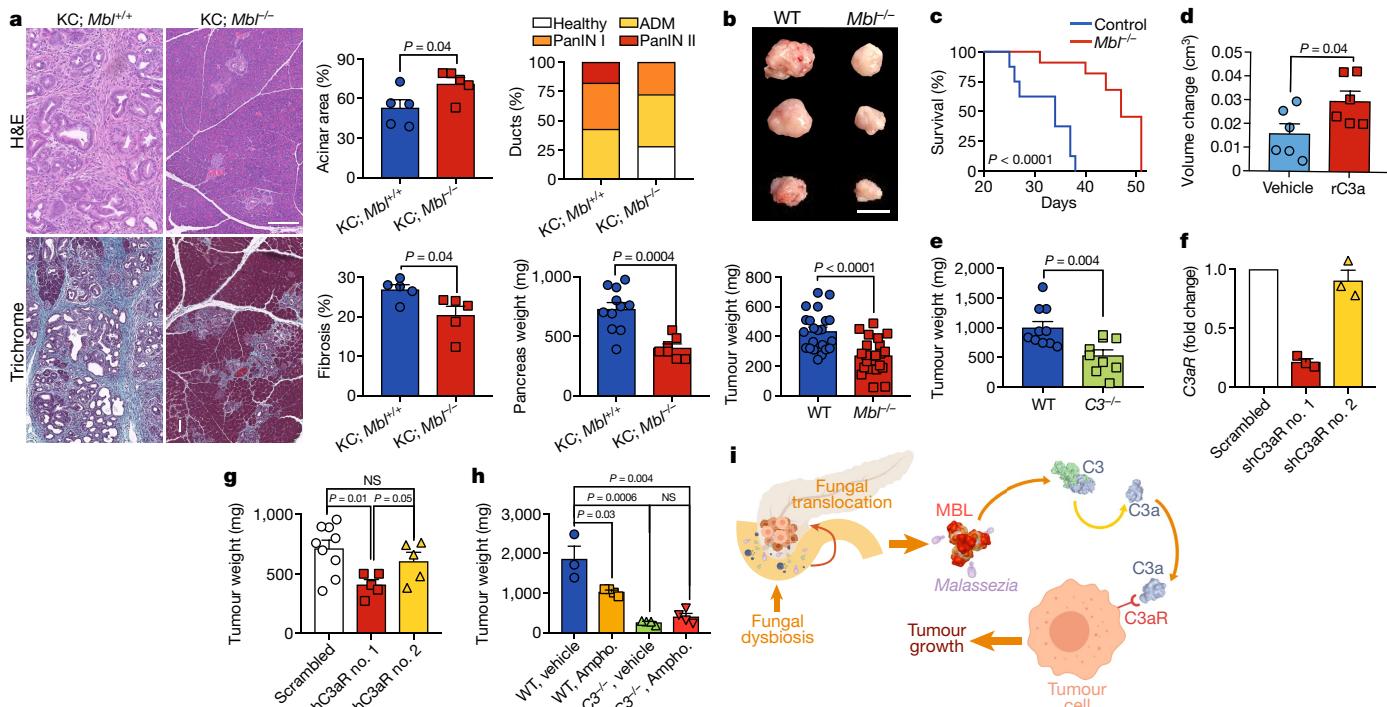
**Fig. 4 | Fungi promote progression of PDA via the MBL–C3 axis. a**, KC (KC; $Mbl^{+/+}$, $n = 11$; used as control) and KC, MBL-null (KC; $Mbl^{-/-}$, $n = 7$) mice were killed at three months old. Tumours were weighed and stained using H&E or trichrome, and analysed for pancreatic dysplasia and fibrosis as in Fig. 3a. Data for control KC mice are the same those shown for the vehicle treatment in Fig. 3. Scale bars, 200 μm. **b, c**, Wild-type and MBL-null mice were administered orthotopic tumour cells from a KPC mouse, and analysed for tumour growth at three weeks ($n = 22$ mice per arm) (**b**) and survival ($n = 8$ wild-type and 5 MBL-null mice) (**c**). Data are representative of at least five experiments. **d**, $Mbl^{-/-}$ host mice were administered orthotopic tumours from KPC mice, received intratumoral injections of recombinant C3a (rC3a) ($n = 6$ mice) or vehicle ($n = 6$ mice) on day 14 via laparotomy, and then volumes were measured. Tumours were collected on day 21, and the change in tumour volume since the injection was calculated. This experiment was repeated twice. **e**, Wild-type ($n = 10$)

and $C3^{-/-}$ ($n = 9$) mice were administered orthotopic tumours from KPC mice, and analysed for tumour growth at three weeks. Data are representative of three experiments. **f, g**, Wild-type mice were orthotopically implanted with tumour cells from KPC mice, treated with short hairpin RNA (shRNA) against $C3aR$ (also known as $C3ar1$) or with control scrambled shRNA. Separate shRNA vectors were used for each treatment. **f**, The efficacy of C3aR knockdown was measured by qPCR ($n = 3$ mice per group). **g**, Quantitative analysis of tumour weights at day 21 ($n = 9$ for scrambled shRNA, and $n = 5$ mice for shC3aR 1 and 2). **h**, Wild-type and $C3^{-/-}$ mice treated with vehicle ($n = 3$ wild-type and 4 $C3^{-/-}$ mice) or amphotericin B ($n = 4$ mice of each background) were administered orthotopic tumours from KPC mice, and killed three weeks later. Tumours were collected and weighed. Data are representative of two experiments. **i**, Schematic depicts the mycobiome–MBL axis in pancreatic oncogenesis. Data are mean ± s.e.m. $P$ values determined by two-tailed Student's $t$-test (**a, b, d–h**), or log-rank test (**c**).

In summary, we found that fungi migrate from the gut to the pancreas, and PDA tumours contain a marked expansion in the pancreatic mycobiome. The composition of the PDA mycobiome was distinct from that of the gut or normal pancreas, and was enriched for *Malassezia* species in both mice and humans. Ablation of the mycobiome was protective against progression of PDA, and repopulation with species of *Malassezia*—but not with other commensal fungi—accelerated oncogenesis. Whether the reprogramming of the mycobiome is a cause or consequence of oncogenesis is difficult to answer fully. However, our fungal adoptive-transfer and fungal-ablation experiments suggest that particular species of fungi are sufficient to promote the progression of PDA. It is likely that—akin to observations regarding the microbiome[1]—inflammation induced by oncogenic *Kras* leads to fungal dysbiosis, which in turn promotes tumour progression via the activation of the MBL–C3 cascade (Fig. 4i). As the mycobiome influences the microbiome and vice versa[8], further study is required to assess this dynamic crosstalk in the pathogenesis of PDA. Our work suggests that the mycobiome may be a new target for therapeutic agents, and an area for the discovery of biomarkers.

## Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and

statements of data and code availability are available at https://doi.org/10.1038/s41586-019-1608-2.

1. Pushkalar, S. et al. The pancreatic cancer microbiome promotes oncogenesis by induction of innate and adaptive immune suppression. *Cancer Discov.* **8**, 403–416 (2018).
2. Hingorani, S. R. et al. Preinvasive and invasive ductal pancreatic cancer and its early detection in the mouse. *Cancer Cell* **4**, 437–450 (2003).
3. Hingorani, S. R. et al. *Trp53^R172H* and *Kras^G12D* cooperate to promote chromosomal instability and widely metastatic pancreatic ductal adenocarcinoma in mice. *Cancer Cell* **7**, 469–483 (2005).
4. van Asbeck, E. C., Hoepelman, A. I., Scharringa, J., Herpers, B. L. & Verhoef, J. Mannose binding lectin plays a crucial role in innate immunity against yeast by enhanced complement activation and enhanced uptake of polymorphonuclear cells. *BMC Microbiol.* **8**, 229 (2008).
5. Ishikawa, T. et al. Identification of distinct ligands for the C-type lectin receptors mincle and dectin-2 in the pathogenic fungus *Malassezia*. *Cell Host Microbe* **13**, 477–488 (2013).
6. Afshar-Kharghan, V. The role of the complement system in cancer. *J. Clin. Invest.* **127**, 780–789 (2017).
7. Cho, M. S. et al. Autocrine effects of tumor-derived complement. *Cell Reports* **6**, 1085–1095 (2014).
8. Sam, Q. H., Chang, M. W. & Chai, L. Y. The fungal mycobiome and its interaction with gut bacteria in the host. *Int. J. Mol. Sci.* **18**, 330 (2017).

## METHODS

No statistical methods were used to predetermine sample size. The experiments were not randomized and investigators were not blinded to allocation during experiments and outcome assessment.

**Mice and tumour models.** KC mice, which develop spontaneous pancreatic neoplasia by targeted expression of mutant *Kras* in the pancreas[2], were a gift from D. Bar-Sagi. C57BL/6, MBL-null and $C3^{-/-}$ mice were originally purchased from Jackson Laboratories and were bred in-house. Littermates were used as controls. Mice were housed in specific-pathogen-free conditions and fed standard mouse chow. In select experiments, C57BL/6 mice generated and housed in a germ-free facility were used. Longitudinal cohort studies were conducted to monitor microbial communities throughout experiments, by serially collecting faecal specimens from littermate wild-type and KC mice. For orthotopic-tumour experiments, 8–10-week-old mice were used. Both male and female mice were used, but mice were sex- and age-matched within each experiment. Mice were administered intra-pancreatic injections of FC1242 tumour cells, derived from the pancreata of KPC mice ($10^5$ cells in Matrigel; BD Biosciences), and killed three weeks after injection. The development of the FC1242 cell line has previously been reported[9]. Cells tested negative for mycoplasma within the past two months. In select experiments, mice were treated with intraperitoneal injection of gemcitabine (1.2 mg twice weekly; MedChemExpress). In other experiments, mice received a single intratumoral injection of recombinant mouse C3a (40 μg/kg; R&D) on day 14 after injections of orthotopic tumours. Mice with pancreatic tumours were monitored regularly for distention of the abdomen, reduced feeding, weight loss, dehydration, hunched posture or poor grooming habits. On detection of signs or symptoms of distress, or when tumour size was estimated by palpation to exceed 15% of the normal body weight, mice were euthanized. Pancreatitis was induced using a regimen of 7 hourly intraperitoneal injections of caerulein (50 μg/kg; Sigma-Aldrich) for 3 consecutive days, before mice were killed 12 h later. Levels of serum amylase activity in mouse serum were measured using the colorimetric mouse amylase assay kit (ab102523, Abcam), according to the manufacturer protocols. Proliferation of KPC tumour cells in vitro was assessed using the 2,3-bis (2-methoxy-4-nitro-5-sulfophenyl)-5-[(phenylamino) carbonyl]-2*H*-tetrazolium hydroxide (XTT) assay (Sigma-Aldrich). Recombinant mC3a (5 nM; R&D) was added to selected wells.

**Antifungal treatment and fungal-transfer experiments.** To ablate the mycobiome in mice, amphotericin B (1 mg/ml; MP Biomedicals) was administered to mice by oral gavage daily for five consecutive days, in addition to adding amphotericin B (0.5 μg/ml) to drinking water for the duration of the experiment[10]. Controls were gavaged with PBS. Orthotopic PDA-tumour cells were administered, or pancreatitis was initiated, three weeks after the start of treatment with amphotericin B. Alternatively, mice were treated with fluconazole (0.5 mg/ml; MP Biomedicals) for three weeks before tumour implantation, using the same regimen[11]. For species-specific repopulation experiments, *M. globosa* (MYA-4612, $1 \times 10^8$ colony-forming units (CFU) per millilitre), *S. cerevisiae* (7752, $1 \times 10^8$ CFU/ml), *C. tropicalis* (MYA-3404, $1 \times 10^8$ CFU/ml; all ATCC), *Candida* sp. (clinical isolate; $1 \times 10^8$ CFU/ml) or *Aspergillus* sp. (clinical isolate; $1 \times 10^6$ CFU/ml) were used to orally gavage mice, after fungal ablation with amphotericin B. Orthotopic PDA cells were administered to recipient mice seven days after repopulation. To assess fungal translocation to the pancreas, $1 \times 10^8$ CFU of GFP-labelled *S. cerevisiae* (ATCC MYA-2011) were introduced via oral gavage, and pancreatic samples were examined at 30 min by flow cytometry. All experiments were approved and in compliance with the New York University School of Medicine Institutional Animal Care and Use Committee.

**C3aR knockdown.** Lentiviral transfer plasmids against C3aR SHCLNG-NM_009779 (TRCN0000027362; CCAGAAAGCAATTCTACTGAT and TRCN0000027385; CCCGTATTTGTATACCGTGAT) were transformed into Stbl3 bacteria. Plasmids were purified using MaxiPrep Kit (Qiagen) and DNA concentration was evaluated by Nanodrop (Thermo Fisher Scientific). The transfer plasmids were co-transfected into HEK293FT cells with packaging plasmids PLP1, PLP2 and VSVG. To evaluate lentivirus concentration, titration of the ability of virus to induce puromycin-resistant colonies was performed in the HEK293FT cell line. Next, KPC tumour cells were transduced for 48 h, followed by selection with puromycin (2 μg/ul) for 10 days. The efficacy of C3aR knockdown was confirmed by qPCR.

**qPCR.** qPCR was performed in duplicate for each sample, using the BioRad Real-Time PCR System (BioRad). Each reaction mixture contained 10 μl of SYBR Green Master Mix (Applied Biosystems), 0.5 μl of forward and reverse primers (Invitrogen) and 3 μl of cDNA (corresponding to 50 ng of RNA). The qPCR conditions were: 50 °C for 2 min, 95 °C for 10 min, followed by 40 cycles at 95 °C for 15 s and 60 °C for 1 min. The amplification of specific transcripts was confirmed by melting-curve profiles, generated at the end of the PCR program. The expression levels of target genes were normalized to the expression of *Gapdh* (internal control) and calculated on the basis of the comparative cycle threshold method ($2^{-\Delta\Delta Ct}$). The *C3aR* primer sequences used in the study were:

forward, TAACCAGATGAGCACCACCA and reverse, TGTGAATGTTGTGTGCATTG.

**Histology, immunohistochemistry and microscopy.** For histological analysis, pancreatic specimens were fixed with 10% buffered formalin, dehydrated in ethanol, embedded with paraffin and stained with H&E or Gomori Trichrome. The percentage of preserved acinar area and fibrosis were calculated, as previously described[1]. The fraction and number of ducts that contained any grade of PanIN lesions were measured by examining 10 H&E-stained high-power fields (40× magnification) per slide. PanINs were graded according to established criteria[12]. In PanIN I ducts, the normal cuboidal pancreatic epithelial cells transition to columnar architecture, and can gain polyploid morphology. PanIN II lesions are associated with a loss of polarity. PanIN III lesions (or in-situ carcinoma) show cribriform morphology, the budding off of cells and luminal necrosis with marked cytological abnormalities, without invasion beyond the basement membrane. The characteristics of control KC mice have previously been detailed[13]. Pancreatic oedema was calculated by measuring intralobular white space on H&E sections. Immunohistochemistry was performed using antibodies directed against CD45 (30-F11, BD Biosciences), C3a (JF10-30, Novus), and DAPI (no. H-1200; Vector Laboratories). For paraffin-embedded slides, samples were dewaxed in ethanol, followed by antigen retrieval with 0.01 M sodium citrate with 0.05% Tween.

**FISH.** The D223 28S rRNA gene probe labelled with the 5′ Cy3 fluorophore (extinction wavelength, 555 nm and emission wavelength, 570 nm; Molecular Probes) was used to detect the fungal colonization within human and mouse pancreatic tissues by FISH. Fluorescence microscopic analysis was conducted with Nikon Eclipse 90i confocal microscope (Nikon) using a Cy3-labelled-probe at 350 pmol/ml, as previously described[1].

**Human sample collection and data from TCGA.** Human faecal samples and specimens of pancreatic tissue were collected under sterile conditions from healthy volunteers and patients undergoing surgery for PDA or for pancreatic endocrine tumours (benign disease) at NYU Langone Medical Center. Donors were de-identified. Samples were stored at −80 °C until analysis. Patients who had received antibiotic or antifungal treatment within the past three months were excluded. Human specimens were collected in compliance with the policies and approval of NYU School of Medicine's Institutional Review Board, and conducted in accordance with the Declaration of Helsinki, the Belmont Report and US Common Rule. Data on gene expression in human tissues was derived from TCGA (https://portal.gdc.cancer.gov/). Survival was measured according to the Kaplan–Meier method, and analysed using the log-rank test.

**Extraction and sequencing of fungal DNA.** Samples of pancreatic tissue were suspended in 500 μl sterile PBS, and pretreated by vortexing and sonication, followed by overnight treatment with proteinase K (2.5 μg/ml; Thermo Fisher) at 55 °C. Total microbial genomic DNA was purified from tissue and faecal samples using the MoBio Power kit, as per the manufacturer's instructions (MoBio Laboratories). DNA was quantified for concentration and purity using the NanoDrop 2000 spectrophotometer (Thermo Fisher) and stored at −20 °C. For the preparation and sequencing of a high-throughput ITS library, the ITS1 region of the 18S rRNA gene was amplified from the genomic DNA of mice or of human faecal samples and samples of pancreatic tissue, according to the modified Illumina metagenomics protocol (part no. 15044223 rev. B). The purified DNA was quantified fluorometrically by Quant-iT PicoGreen assay (Molecular Probes) in a SpectraMax M5 microplate reader (Molecular Devices), and the concentration was adjusted to 10 ng/μl for all sequencing assays. PCR was initially performed using the primer set ITS1F (5′-CTTGGTCATTTAGAGGAAGTAA-3′) and ITS2 (5′-GCTGCGTTCTTCATCGATGC-3′)[14]; each with overhang adaptor sequences (IDT) using 2× Kapa HiFi Hotstart ReadyMix DNA polymerase (KapaBiosystems). Samples were amplified in duplicates and purified using AMPure XP beads. Amplification was performed at 95 °C (5 min), with 25 cycles of 95 °C (1 min), 53 °C (45 s), 72 °C (1 min) and a final extension of 72 °C (10 min). Dual indices from Illumina Nextera XT index kits (Illumina) were added to target amplicons in a second PCR using the 2× Kapa HiFi Hotstart ReadyMix DNA polymerase. PCR conditions were 95 °C (5 min), with 10 cycles of 95 °C (1 min), 53 °C (45 s), 72 °C (1 min) and a final extension of 72 °C (10 min). After each PCR cycle, purified libraries of AMPure XP beads were checked for purity by Nanodrop, quantified by PicoGreen assay and sizes were confirmed on agarose gels. Negative controls were included in all sequencing runs. Equimolar amounts of the generated libraries were combined and quantified fluorometrically. The pooled amplicon library was denatured, diluted and sequenced on an Illumina MiSeq platform using MiSeq Reagent Kit v.3 (600 cycles) following the 2× 300-bp paired-end sequencing protocol.

**Bioinformatics and statistical analyses.** The Illumina-generated fungal ITS sequence data were processed using QIIME (v.1.9.1), and the reads were demultiplexed, quality-filtered and clustered into OTUs using default parameters[15]. To maintain consistency, read 1 was used for the analyses, as previously described[14]. Before demultiplexing, the 5′ primers of a total 16,647,630 R1 reads were trimmed

using cutadapt (v.1.12), and sequences that were shorter than 100 bases or sequences including asparagine were discarded. The reads were filtered by quality at 20, using multiple_split_libraries_fastq.py ($q = 19$; defaults were used for the other parameters). The 1,989,618 quality reads (mean 8,575; $n = 166$) were then processed with QIIME. Chimeric sequences were removed using VSEARCH (v.2.4.3) with UNITE UCHIME reference dataset (v.7.2). OTUs were picked using the open-reference OUT picking method, with default parameters, against the UNITE reference database (v.7.2) to assign taxonomy using pick_open_reference_otus.py[16]. There were 126,862 OTUs, corresponding to 1,856,993 reads (about 93.57% of the total reads), that did not align to fungi; these OTUs were excluded from the downstream analyses. OTUs that were unidentified in UNITE database were blasted to NCBI, and the taxonomy information of the best hit (similarity or coverage $\geq$ 97%) for each OTU was re-assigned. A total of 127,646 sequence reads were clustered into 1,899 OTUs (corresponding to 86,640 reads) for longitudinal faecal samples from mice; 390 OTUs (corresponding to 25,021 reads) for tissue samples from mice; 2,980 OTUs (corresponding to 15,349 reads) for faecal samples from humans; and 311 OTUs (corresponding to 636 reads) for tissue samples from humans. Sequence data were analysed at various levels of phylogenetic affiliations. Low-abundance OTUs in <2 samples, and samples identified as outliers, were removed. Distinctions in the composition of the mycobiomes between cohorts and within samples over time were tested for significance using a Mann–Whitney $U$ test. Alpha-diversity and beta-diversity were computed and plotted in Phyloseq. PCoA was performed on Bray–Curtis dissimilarity indices, and a one-way PERMANOVA was used to test for significant differences between cohorts (Adonis, R package Vegan v.2.4.5). $P$ values < 0.05 were considered to be significant.

**Quality control.** For quality control, we used best practices for microbiome- and mycobiome-based studies, as previously described[1]. All the samples were collected using sterile techniques. All PCR reagents were regularly checked for environmental contaminants using ITS universal primers. All qPCR reactions had appropriate controls (without template) to exclude DNA contaminants. To control for the quality of our sequencing, we used both predetermined mock communities (such as *C. tropicalis*) and 'negative' (reagent-only) controls, to check background contamination and the rate of sequencing errors. We included both of these controls in each of the sequencing runs. We further confirmed the quality of our sequencing by including community controls composed of predetermined ratios of DNA from a mixture of three fungal species.

**Figure preparation.** Figures were prepared using BioRender software and Indesign (Adobe).

**Reporting summary.** Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

## Data availability

The sequence datasets analysed in this article are publicly available in the NCBI BioProject database, under the accession number PRJNA557226. Raw data for all experiments are available as Source Data to the relevant figures. Any other relevant data are available from the corresponding authors upon reasonable request.

9. Zambirinis, C. P. et al. TLR9 ligation in pancreatic stellate cells promotes tumorigenesis. *J. Exp. Med.* **212**, 2077–2094 (2015).
10. Reikvam, D. H. et al. Depletion of murine intestinal microbiota: effects on gut mucosa and epithelial gene expression. *PLoS ONE* **6**, e17996 (2011).
11. Skalski, J. H. et al. Expansion of commensal fungus *Wallemia mellicola* in the gastrointestinal mycobiota enhances the severity of allergic airway disease in mice. *PLoS Pathog.* **14**, e1007260 (2018).
12. Hruban, R. H. et al. Pancreatic intraepithelial neoplasia: a new nomenclature and classification system for pancreatic duct lesions. *Am. J. Surg. Pathol.* **25**, 579–586 (2001).
13. Seifert, L. et al. The necrosome promotes pancreatic oncogenesis via CXCL1 and Mincle-induced immune suppression. *Nature* **532**, 245–249 (2016).
14. Walters, W. et al. Improved bacterial 16S rRNA gene (V4 and V4-5) and fungal internal transcribed spacer marker gene primers for microbial community surveys. *mSystems* **1**, e00009-15 (2015).
15. Navas-Molina, J. A. et al. Advancing our understanding of the human microbiome using QIIME. *Methods Enzymol.* **531**, 371–444 (2013).
16. Caporaso, J. G. et al. QIIME allows analysis of high-throughput community sequencing data. *Nat. Methods* **7**, 335–336 (2010).

**Author contributions** B.A. carried out in vivo and in vitro experiments, study design, PCR, analysis and interpretation, manuscript preparation and statistical analysis; S.P. carried out fungal DNA sequencing, analysis and interpretation, manuscript preparation, microbiology study design and statistical analysis; R.C. carried out in vivo experiments, histological analysis and manuscript preparation; Q.L. performed computational analyses and provided critical review; R.A. carried out in vivo experiments and provided technical assistance; J.I.K. carried out in vivo experiments and provided critical review; S.A.S. carried out mouse breeding and histology; D.W. performed tissue culture and cell-line generation; P.P. provided technical assistance and carried out in vivo experiments; N.V. carried out knockdown experiments; Y.G. carried out PCR; A.S. performed FISH and provided critical review; M.V. carried out DNA extraction and contributed to computational analysis; B.D. carried out in vivo experiments and critical review; W.W. provided technical assistance; J.L. provided critical review and contributed to study design; E.K. carried out in vivo experiments and contributed to study design; J.A.K.R. provided technical assistance and contributed to study design; M.H. carried out in vivo experiments; C.Z. carried out human-sample collection; X.L. provided technical assistance; D.S. and G.M. conceived, designed, supervised, analysed and interpreted the study, prepared the manuscript and provided critical review.
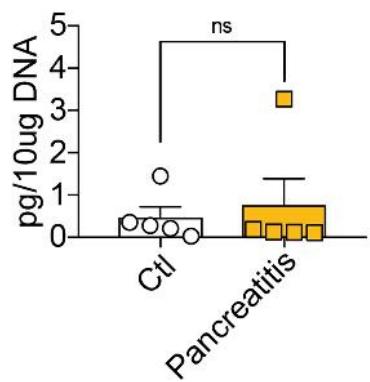
**Extended Data Fig. 1 | Fungal infiltration of the pancreas in benign disease.** Fungal DNA content was tested using qPCR in pancreata from control (ctl) mice ($n = 5$) and mice induced to develop caerulein-induced pancreatitis ($n = 5$). ns, not significant. Data are mean ± s.e.m. Two-tailed Student's $t$-test.
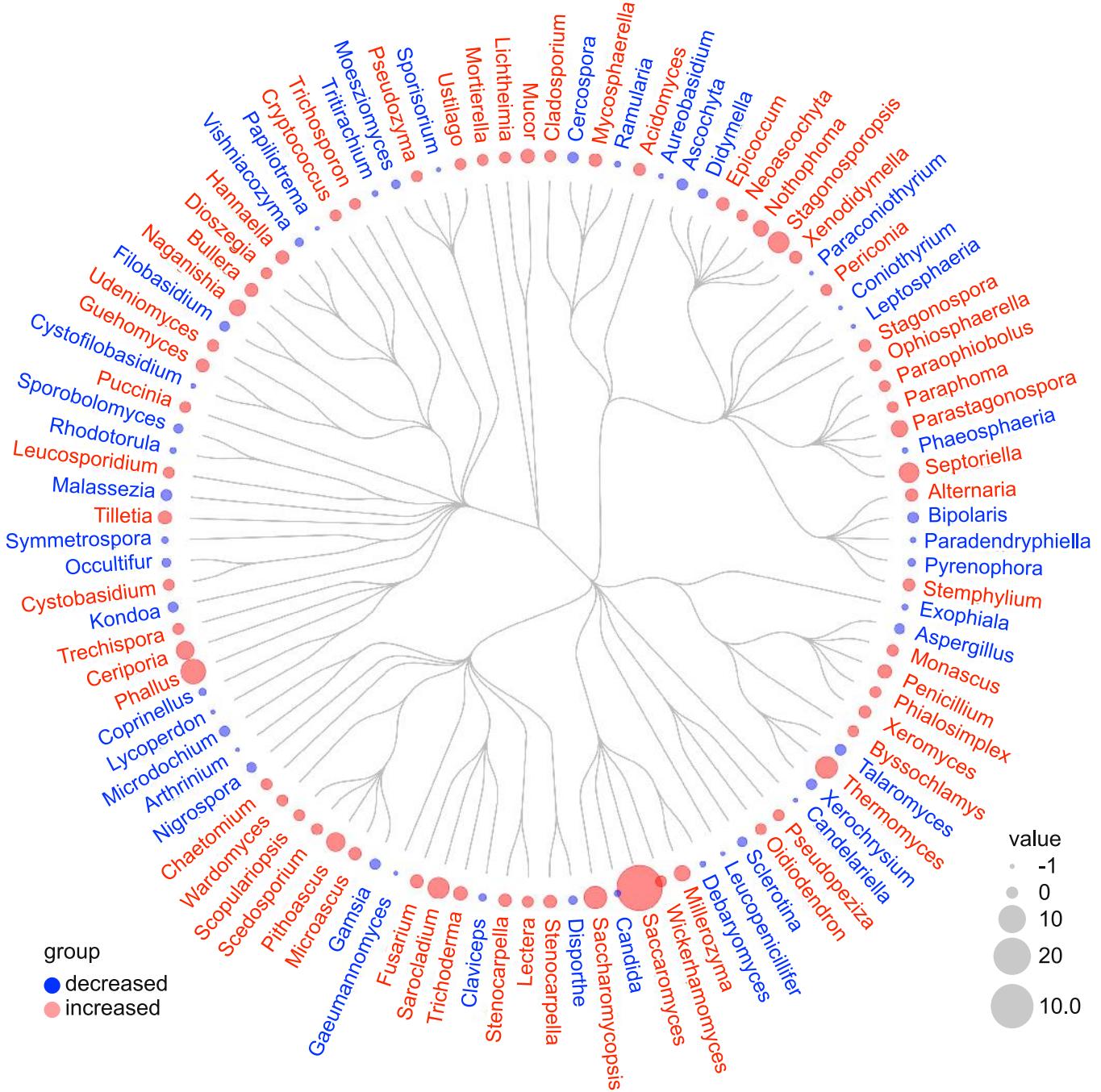
**Extended Data Fig. 2 | Dysbiosis of the gut mycobiome in a mouse model of PDA.** Hierarchical tree cladogram depicting changes in the taxonomic composition of the mycobiome (assigned to the genus level) in the guts of 30-week-old KC ($n = 14$) compared to wild-type ($n = 12$) mice, based on the average percentage relative abundance of genera as determined by 18S ITS sequencing.
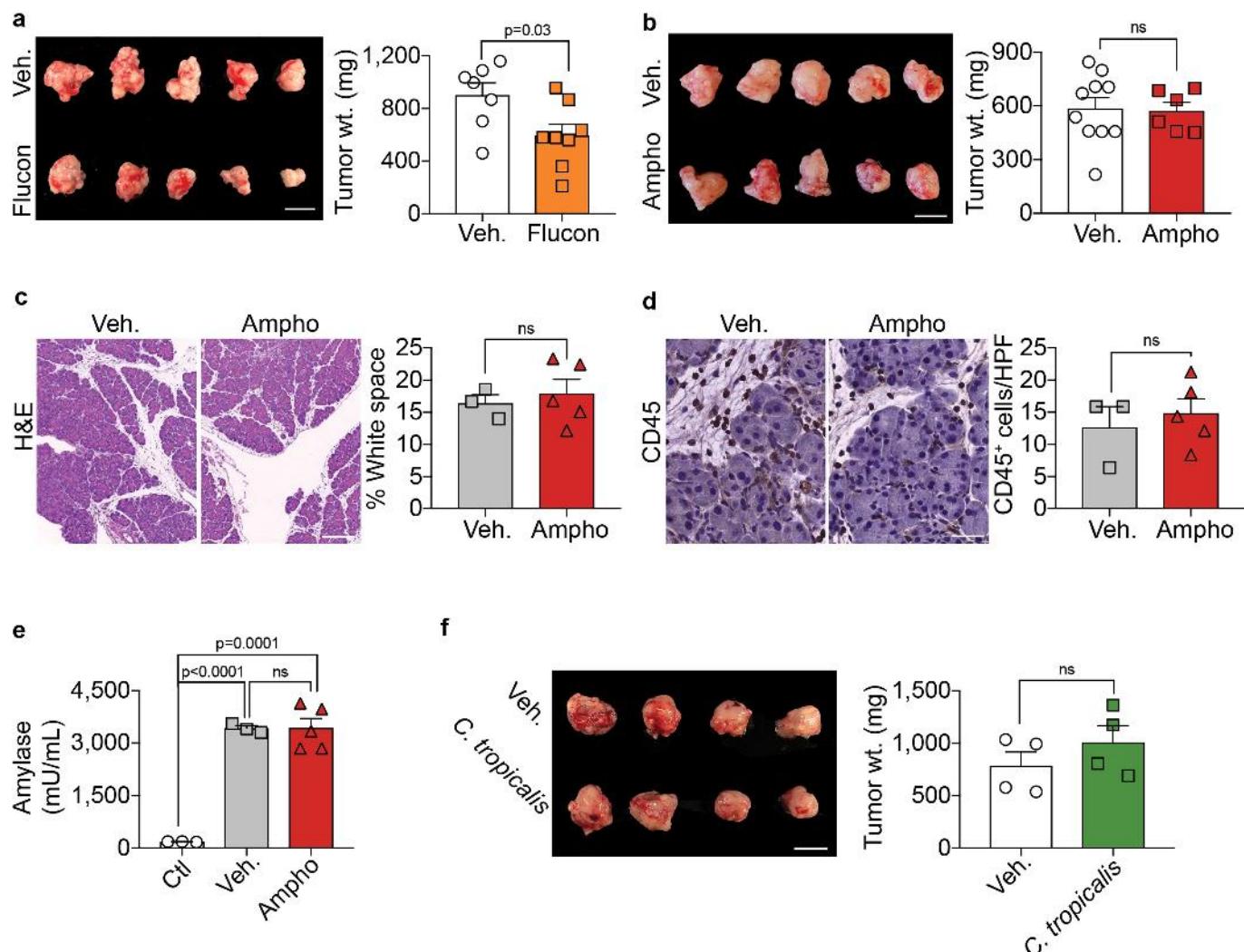
**Extended Data Fig. 3 | Efficacy of antifungal treatments in pancreatic disease. a**, Wild-type mice that bear orthotopic PDA tumours were treated with vehicle ($n = 7$ mice) or fluconazole ($n = 8$ mice), and killed three weeks later. Tumours were collected and weighed. Data are representative of experiments that were performed twice. Scale bar, 1 cm. **b**, Germ-free wild-type mice were treated with amphotericin B ($n = 6$ mice) or vehicle ($n = 10$ mice), and orthotopic tumours from KPC mice were administered to them. Mice were killed three weeks later, and tumours were collected and weighed. Scale bar, 1 cm. **c–e**, Wild-type mice induced to develop caerulein-induced pancreatitis were serially treated with amphotericin B ($n = 5$ mice) or vehicle ($n = 3$ mice). **c**, Representative H&E-stained sections of pancreata are shown, and pancreatic oedema was quantified by measuring the percentage of the area that was white space. Scale bar, 100 μm. **d**, CD45$^+$ inflammatory cell infiltration was determined by immunohistochemistry. Scale bar, 20 μm. **e**, Serum levels of amylase were measured. $n = 5$ mice treated with amphotericin B, $n = 3$ mice treated with vehicle and $n = 3$ mock-treated (control) mice. **f**, Wild-type mice treated with amphotericin B were repopulated with *C. tropicalis* ($n = 4$ mice) or vehicle ($n = 4$ mice), and killed three weeks later. Tumours were collected and weighed. Scale bar, 1 cm. Data are mean ± s.e.m. *P* values determined by two-tailed Student's *t*-test (**a–f**).
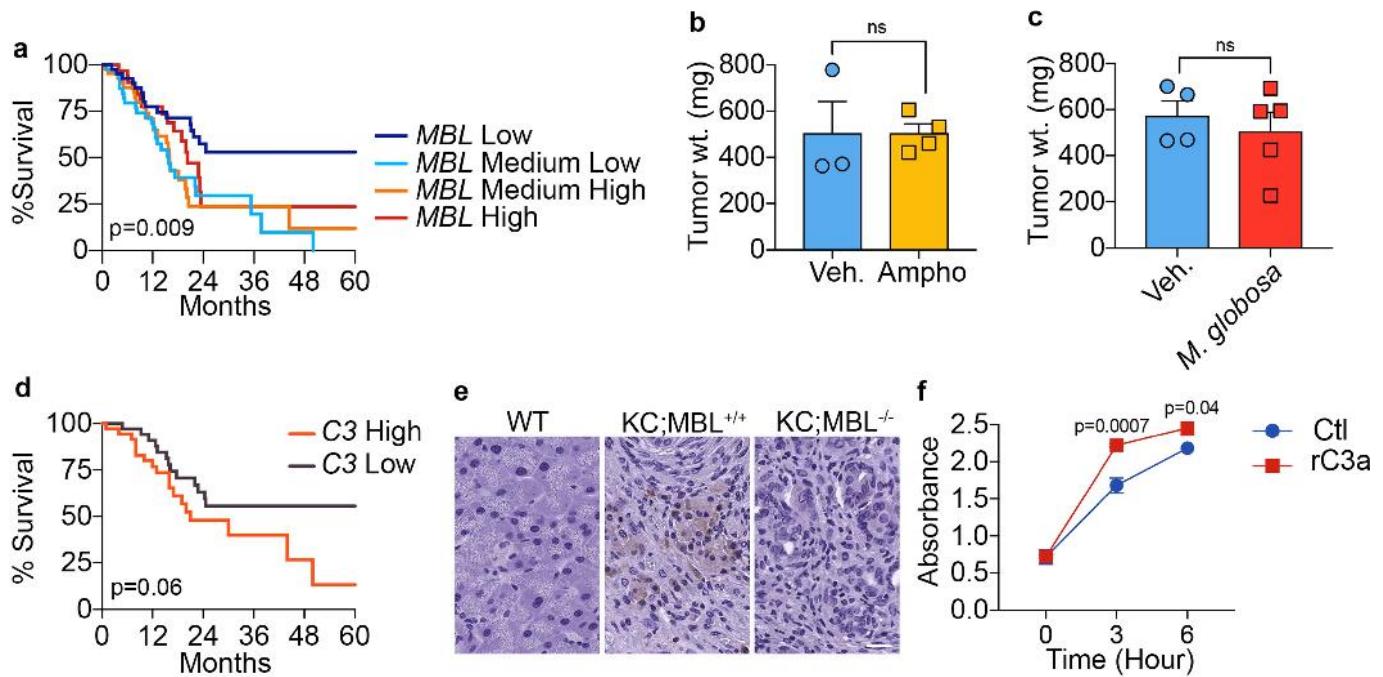
**Extended Data Fig. 4 | Fungal dysbiosis drives the progression of PDA via the lectin pathway. a**, Kaplan–Meier survival curve of patients with PDA, stratified by high ($n = 16$ patients), medium-high ($n = 24$ patients), medium-low ($n = 26$ patients) and low ($n = 17$ patients) expression of *MBL* on the basis of data from TCGA. **b**, Orthotopic tumours from KPC mice were administered to MBL-null mice treated with vehicle ($n = 3$ mice) or amphotericin B ($n = 4$ mice), and killed three weeks later. Tumours were collected and weighed. Data are representative of three separate experiments. **c**, MBL-null mice treated with amphotericin B were repopulated with *M. globosa* ($n = 5$ mice) or sham-repopulated ($n = 4$ mice), and killed three weeks later. Tumours were collected and weighed. Data are representative of experiments that were repeated twice.

**d**, Kaplan–Meier survival curve of patients with PDA, stratified by high ($n = 18$) versus low ($n = 15$) expression of *C3*, on the basis of data from TCGA. **e**, Pancreata from three-month-old wild-type, KC and KC, MBL-null mice were stained using a monoclonal antibody against C3a. Representative images from two experiments are shown. Scale bar, 20 μm. **f**, KPC tumour cells were seeded in 96-well plates with vehicle or recombinant mouse C3a. $n = 5$ cells per group for each time point. Cellular proliferation was measured at serial time points using the XTT assay. Data are representative of experiments that were repeated three times. Data are mean ± s.e.m. *P* values determined by two-tailed log-rank test (**a**, **d**) or two-tailed Student's *t*-test (**b**, **c**, **f**).

| | |
|---|---|
| Corresponding author(s): | George Miller<br>Deepak Saxena |
| Last updated by author(s): | Jul 29, 2019 |

# Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist .

## Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

| n/a | Confirmed | |
|---|---|---|
| ☐ | ✗ | The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement |
| ☐ | ✗ | A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly |
| ☐ | ✗ | The statistical test(s) used AND whether they are one- or two-sided<br>*Only common tests should be described solely by name; describe more complex techniques in the Methods section.* |
| ☐ | ✗ | A description of all covariates tested |
| ☐ | ✗ | A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons |
| ☐ | ✗ | A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals) |
| ☐ | ✗ | For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted<br>*Give P values as exact values whenever suitable.* |
| ☐ | ✗ | For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings |
| ☐ | ✗ | For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes |
| ✗ | ☐ | Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated |

*Our web collection on statistics for biologists contains articles on many of the points above.*

## Software and code

Policy information about availability of computer code

| Data collection | BioRad CFX Manager 3.1, SoftMaxPro 5.4.6, Microsoft Excel 2016, SoftMaxPro 5.4.6, BioRad CFX Manager 3.1, QIIME 1.9.1, Nikon i Series Support Tools Ver2.4.4, Illumina |
|---|---|
| Data analysis | GraphPad Prism 7, Microsoft Excel 2016, BioRad CFX Manager 3.1, SoftMaxPro 5.4.6, R 3.5.2, LefSe, cutadapt v.1.12, VSEARCH v.2.4.3, Phyloseq 1.27.0, Adonis, Vegan v.2.4.5, igraph 1.2.4, R 3.5.2, UNITE database v. 7.2, QIIME 1.9.1 |

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

## Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Sequence data will be available in the Sequence Read Archive (SRA) database at NCBI at the time of publication. All other datasets generated during and/or analysed during the current study are available from the corresponding author on reasonable request.

# Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

# Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

| | |
|---|---|
| Sample size | Power analyses were done based on estimates |
| Data exclusions | Data points were not excluded. |
| Replication | All experiments were repeated multiple times as indicated in each figure legend. Attempts at replication were successful. |
| Randomization | No formal randomization was carried out in experiments involving mutliple genotypes. For all other experiments, animals were randomly divided into experimental groups. |
| Blinding | Administration of compounds was carried out as a blinded experiment (all information about the expected outputs and the nature of used compounds were kept from the animal-technicians). |

# Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

## Materials & experimental systems

| n/a | Involved in the study |
|---|---|
| ☐ | ✗ Antibodies |
| ☐ | ✗ Eukaryotic cell lines |
| ✗ | ☐ Palaeontology |
| ☐ | ✗ Animals and other organisms |
| ☐ | ✗ Human research participants |
| ✗ | ☐ Clinical data |

## Methods

| n/a | Involved in the study |
|---|---|
| ✗ | ☐ ChIP-seq |
| ✗ | ☐ Flow cytometry |
| ✗ | ☐ MRI-based neuroimaging |

## Antibodies

| | |
|---|---|
| Antibodies used | rat anti CD45 (clone 30-F11, BD Biosciences) catalog number 553076 (1:100); rabbit anti C3a (clone JF10-30, NOVUS) catalog number NBP2-66994 (1:100). Secondary antibodies: Rat IgG HRP-conjugated Antibody (Vector Labs) catalog number MP-7404 (1:1000); Rabbit IgG HRP-conjugated Antibody (Vector Labs) catalog number MP-7401 (1:1000). |
| Validation | Antibody specificity was evaluated using the proper negative controls (rat IgG2b, κ for rat anti CD45 and rabbit IgG for anti C3a). |

## Eukaryotic cell lines

Policy information about cell lines

| | |
|---|---|
| Cell line source(s) | Cells used for orthotopic tumor injections were generated by our group from endogenous tumors. HEK293FT were purchased from Thermo Fisher Scientific (catalog number: R70007). |
| Authentication | None of the cell lines were authenticated. |
| Mycoplasma contamination | Mycoplasma testing was performed within the past 2 months and was negative |
| Commonly misidentified lines (See ICLAC register) | No commonly misidentified cell lines were used. |

## Animals and other organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research

| | |
|---|---|
| Laboratory animals | C57BL/6, MBL-null (B6.129S4-Mbl1tm1Kata Mbl2tm1Kata/J), and C3–/– (B6;129S4-C3tm1Crr/J) mice were originally purchased from Jackson Labs (Bar Harbor, ME). 8-10 week old males and females were used. KC mice, which develop spontaneous pancreatic neoplasia by targeted expression of mutant Kras in the pancreas, were a gift from Dafna Bar-Sagi (New York University). |
| Wild animals | This study did not involve wild animals. |

| Field-collected samples | This study did not involve samples collected from the field. |
| --- | --- |
| Ethics oversight | All animal experiments were approved by the New York University School of Medicine Institutional Animal Care and Use Committee (IACUC). |

Note that full information on the approval of the study protocol must also be provided in the manuscript.

# Human research participants

Policy information about [studies involving human research participants](#)

| Population characteristics | Human fecal samples and pancreatic tissue specimens were sterilely collected from healthy volunteers and patients undergoing surgery for either PDA or benign disease (pancreatic endocrine tumors) at NYU Langone Medical Center. |
| --- | --- |
| Recruitment | Human fecal samples and pancreatic tissue specimens were sterilely collected from healthy volunteers and patients undergoing surgery for either PDA or benign disease (pancreatic endocrine tumors) at NYU Langone Medical Center. |
| Ethics oversight | Human specimens were obtained using an Institutional Review Board approved protocol, conducted in accordance with the Declaration of Helsinki, the Belmont Report, and U.S. Common Rule and donors of de-identified specimens gave informed consent. |

Note that full information on the approval of the study protocol must also be provided in the manuscript.

# LETTER

# Inducing and exploiting vulnerabilities for the treatment of liver cancer

Cun Wang[1,2,7], Serena Vegna[3,7], Haojie Jin[1,2,7], Bente Benedict[3,7], Cor Lieftink[2], Christel Ramirez[3], Rodrigo Leite de Oliveira[2], Ben Morris[2], Jules Gadiot[3], Wei Wang[4], Aimée du Chatinier[2], Liqin Wang[2], Dongmei Gao[5], Bastiaan Evers[2], Guangzhi Jin[6], Zheng Xue[2], Arnout Schepers[2], Fleur Jochems[2], Antonio Mulero Sanchez[2], Sara Mainardi[2], Hein te Riele[3], Roderick L. Beijersbergen[2], Wenxin Qin[1]*, Leila Akkari[3]* & René Bernards[2]*

**Liver cancer remains difficult to treat, owing to a paucity of drugs that target critical dependencies[1,2]; broad-spectrum kinase inhibitors such as sorafenib provide only a modest benefit to patients with hepatocellular carcinoma[3]. The induction of senescence may represent a strategy for the treatment of cancer, especially when combined with a second drug that selectively eliminates senescent cancer cells (senolysis)[4,5]. Here, using a kinome-focused genetic screen, we show that pharmacological inhibition of the DNA-replication kinase CDC7 induces senescence selectively in liver cancer cells with mutations in *TP53*. A follow-up chemical screen identified the antidepressant sertraline as an agent that kills hepatocellular carcinoma cells that have been rendered senescent by inhibition of CDC7. Sertraline suppressed mTOR signalling, and selective drugs that target this pathway were highly effective in causing the apoptotic cell death of hepatocellular carcinoma cells treated with a CDC7 inhibitor. The feedback reactivation of mTOR signalling after its inhibition[6] is blocked in cells that have been treated with a CDC7 inhibitor, which leads to the sustained inhibition of mTOR and cell death. Using multiple in vivo mouse models of liver cancer, we show that treatment with combined inhibition of of CDC7 and mTOR results in a marked reduction of tumour growth. Our data indicate that exploiting an induced vulnerability could be an effective treatment for liver cancer.**

The increase in the incidence of hepatocellular carcinoma (HCC)[2], the fact that HCC mutations cannot currently be treated with drugs, and unresponsiveness of these tumours to therapy highlight the urgent need to develop new therapeutic approaches for treating this disease[7]. A 'one–two punch' approach to cancer therapy has previously been proposed, in which the first drug induces a vulnerability that is exploited by the second drug[4,5]. Senescent cells have distinct cellular features, which can confer sensitivity to senolytic agents[8,9]. Here we experimentally validate the one–two punch therapy for treatment of liver cancer with mutations in *TP53*.

To identify genes that, when inactivated, can induce senescence in liver cancer cells, we used a CRISPR–Cas9 genetic screen with a lentiviral guide RNA (gRNA) library that represents all human kinases in Hep3B and Huh7 liver cancer cells[10] (Fig. 1a). We identified 38 genes that are required for proliferation (Fig. 1b, Extended Data Fig. 1a, Supplementary Table 1), 14 of which could be inhibited by small-molecule compounds (Fig. 1b). We screened compounds that target these 14 kinases for their ability to induce senescence selectively in liver cancer cells (Fig. 1c). XL413—a potent inhibitor of the DNA-replication kinase CDC7[11]—was the most selective of these compounds in inducing senescence-associated β-galactosidase (SA-β-gal, a marker of senescence) in Hep3B and Huh7 cells, as compared to non-transformed BJ and RPE-1 human cell lines (Fig. 1c). These findings suggest that the

inhibition of CDC7 could represent a senescence-inducing strategy in liver cancer.

As seen in several types of tumour[12], liver cancer cell lines express levels of CDC7 that are higher than those of non-transformed cells (Extended Data Fig. 1b). Expression of *CDC7* is upregulated in tumour tissues relative to paired non-tumour tissues in two cohorts of patients with liver cancer ($n = 213$ and $n = 50$) (Fig. 1d), and this was confirmed at the protein level (Extended Data Fig. 1c). Moreover, in a cohort of 365 patients with HCC[13], patients with the highest levels of *CDC7* mRNA in their tumours exhibited the worst survival (Extended Data Fig. 1d).

We treated a panel of non-transformed cells and liver cancer cell lines with increasing concentrations of XL413. Proliferation was impaired in liver cancer cell lines with mutations in *TP53*, whereas liver cancer cell lines with wild-type *TP53* (SK-Hep1 and Huh6 cells)—as well as all four non-transformed cell lines—displayed no sensitivity to XL413 (Fig. 2a). The XL413-sensitive cell line HepG2 is an outlier in this respect but carries a mutation in *ATM*, which acts upstream of p53 in the DNA-damage response. Importantly, knockdown of *TP53* mediated by short hairpin RNA (shRNA) in wild-type cells sensitized these cells to CDC7 inhibition (Extended Data Fig. 1e–g), which indicates a causal relationship between *TP53* mutation status and sensitivity to inhibition of CDC7.

The anti-proliferative effect of XL413 was associated with the induction of senescence markers in liver cancer cells with *TP53* mutations, but not in liver cancer cells with wild-type *TP53* or in non-transformed cells (Fig. 2b, Extended Data Fig. 2a), and a senescence signature[14] was enriched in HCC cells with mutations in *TP53* treated with XL413 (Fig. 2c). The notion that CDC7 inhibition induces a senescence-like state in liver cancer cells with mutations in *TP53* is further supported by the findings that (i) withdrawal of XL413 does not lead to re-entry into the cell cycle in the majority of HCC cells, (ii) treatment with XL413 induced senescence-associated heterochromatin foci and (iii) treatment with XL413 induced expression of a number of cytokines, as part of the senescence-associated secretory phenotype[15] (Extended Data Fig. 2b–d). There was no evidence for substantial induction of apoptosis in HCC cells with *TP53* mutations treated with XL413 (Extended data Fig. 2e). Comparable results were obtained with two CDC7 inhibitors that are unrelated to XL413—LY3177833 and TAK-931 (Extended Data Fig. 3a–f). Consistent with this, knockdown of *CDC7* impaired proliferation and induced senescence in liver cancer cells with *TP53* mutations, but had no effect on cells with wild-type *TP53* (Extended Data Fig. 3g–i).

The phosphorylation of MCM2 (a target of CDC7[12]) was suppressed equally by each of the three CDC7 inhibitors in both wild-type cells and cells with *TP53* mutations (Fig. 2d, Extended Data Fig. 4a, b),

[1]State Key Laboratory of Oncogenes and Related Genes, Shanghai Cancer Institute, Renji Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai, China. [2]Oncode Institute, Division of Molecular Carcinogenesis, The Netherlands Cancer Institute, Amsterdam, The Netherlands. [3]Oncode Institute, Division of Tumour Biology and Immunology, The Netherlands Cancer Institute, Amsterdam, The Netherlands. [4]Division of Cell Biology, The Netherlands Cancer Institute, Amsterdam, The Netherlands. [5]Liver Cancer Institute, Zhongshan Hospital, Fudan University, Key Laboratory of Carcinogenesis and Cancer Invasion, Ministry of Education, Shanghai, China. [6]Department of Pathology, Eastern Hepatobiliary Surgery Hospital, Second Military Medical University, Shanghai, China. [7]These authors contributed equally: Cun Wang, Serena Vegna, Haojie Jin, Bente Benedict. *e-mail: wxqin@sjtu.edu.cn; l.akkari@nki.nl; r.bernards@nki.nl
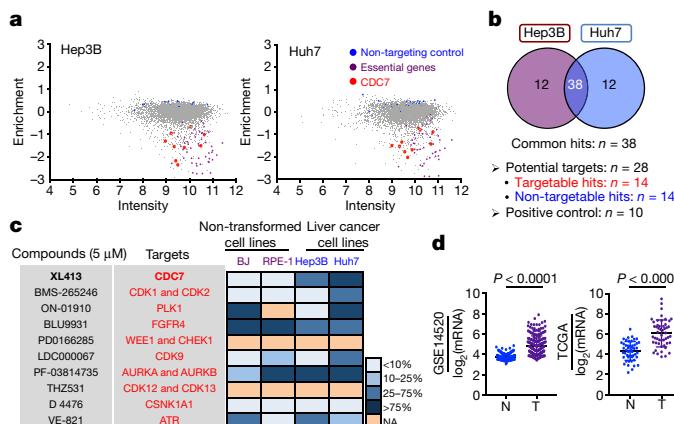
**Fig. 1 | A two-step screen identifies CDC7 as a target for the senescence-inducing strategy in liver cancer. a**, Hep3B and Huh7 cells were transduced with a lentiviral kinome gRNA library and 3 independent replicates were cultured for 14 days. gRNA barcodes from samples at day 0 and day 14 were recovered by PCR, and analysed by next-generation sequencing. The $y$ axis shows $\log_2$ scale in abundance (ratio of gRNA frequency in the day-14 sample to that in the day-0 sample). The $x$ axis shows $\log_2$ scale of the average read-count in the samples of day 0. **b**, Thirty-eight common hits (among the top 50 most-strongly depleted hits in each cell line) were identified by CRISPR screen in Hep3B and Huh7 cells. **c**, Heat map indicates the effects of compounds (10 compounds targeting 14 hits identified by CRISPR screen, 5 μM for 4-day treatments) on inducing SA-β-gal activity in non-transformed cell lines (BJ and RPE-1) and liver cancer cell lines (Hep3B and Huh7). NA, not applicable. **d**, *CDC7* mRNA expression in paired tumour (T) and non-tumour (N) tissues from GSE14520 cohort ($n = 213$ patients) and The Cancer Genome Atlas (TCGA) database ($n = 50$ patients). Paired two-sided *t*-test. Data in graphs are mean ± s.d. Data in **a**–**c** are representative of three independent biological experiments.

which indicates that there is no correlation between the cell fate that is induced by CDC7 inhibitors and the degree of inhibition of the downstream targets of CDC7. To further address why it is that inhibition of CDC7 selectively induces senescence in the context of *TP53* mutation, we assessed protein expression associated with DNA damage following treatment with XL413. The induction of γH2AX and DNA double-strand breaks was notable after CDC7 inhibition in liver cancer cells with *TP53* mutations as compared to cells with wild-type *TP53*; these latter instead displayed a clear upregulation of p21 (CIP1) (Fig. 2d, e, Extended Data Fig. 4a–c). This differential effect is most readily explained by the finding that multiple gene signatures associated with DNA repair are upregulated in cells with wild-type *TP53* (SK-Hep1 and BJ) treated with XL413, but are suppressed in cells with *TP53* mutations upon inhibition of CDC7 (Fig. 2f, Extended Data Fig. 4d, e). Consistently, the inhibition of DNA repair with the ATR inhibitor AZD6738, or with the CHK1 inhibitor MK-8776, in liver cancer cells with wild-type *TP53* resulted in increased double-strand breaks when combined with treatment with XL413 (Extended Data Fig. 4f). Inhibition of CDC7 also resulted in a significant increase in the duration of mitosis (Extended Data Fig. 4g, h). We further confirmed the specificity of the effects of CDC7 inhibition in *Trp53*^−/− mouse cell models of liver cancer[16] (Extended Data Fig. 4i, j). Moreover, XL413 induced senescence in non-small-cell lung-cancer cells with *TP53* mutations, but not in cells with wild-type *TP53* (Extended Data Fig. 5a, b). Similarly, in isogenic *TP53*^−/− and *TP53*^+/+ HCT116 colon-cancer cells, the inhibition of CDC7 induced senescence only in *TP53*^−/− cells (Extended Data Fig. 5c–e).

The induction of senescence represents a double-edged sword for tumour control[15,17], and the potentially harmful properties of senescent tumour cells make their elimination therapeutically relevant. The high concentration of ABT263[8], a senolytic BH3 mimetic drug, that is required to promote apoptosis of XL413-induced senescent cells—and the lack of sensitivity of these cells to dasatinib[9]—prevent their
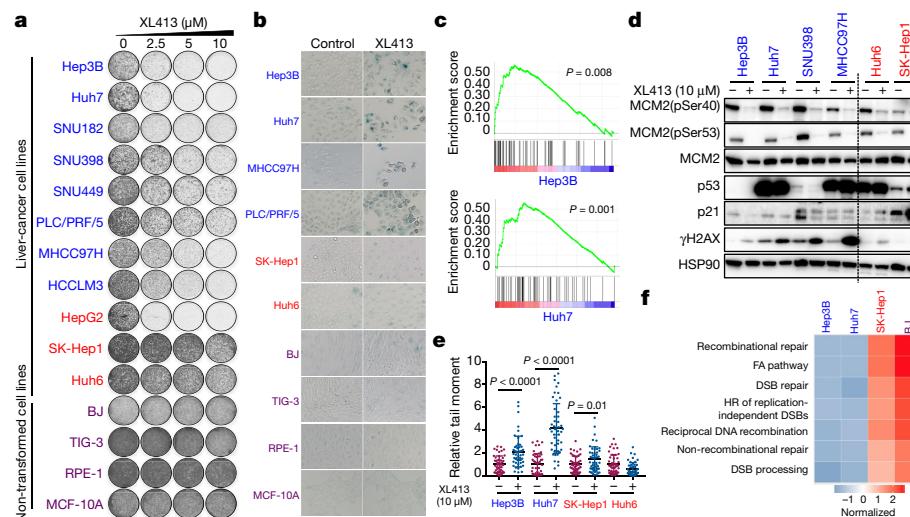


**Fig. 2 | Inhibition of CDC7 selectively induces senescence in liver cancer cells with *TP53* mutations. a**, Long-term colony formation assay was performed over 10–14 days on liver cancer cell lines with *TP53* mutations (blue), liver cancer cell lines with wild-type *TP53* (red) and non-transformed cell lines (purple), cultured with the indicated concentrations of XL413. **b**, Liver cancer cells and non-transformed cell lines were cultured in the presence of 10 μM XL413 for 4 days and the induction of senescence was assessed by staining for SA-β-gal activity. **c**, GSEA of a previously published[14] signature of genes that are upregulated in senescence, in sequencing data from Hep3B and Huh7 cells treated with 10 μM XL413 for 4 days (Methods). **d**, Proteins were extracted from liver cancer cell lines with *TP53* mutations or wild-type *TP53*, treated with 10 μM XL413 for 7 days and then analysed by western blotting.

MCM2(pSer40), MCM2 phosphorylated at Ser40; MCM2(pSer53), MCM2 phosphorylated at Ser53. **e**, Neutral comet assays were performed on Hep3B, Huh7, SK-Hep1 and Huh6 cells cultured with 10 μM XL413 for 7 days. The value of tail moments in each treatment group were normalized on the basis of the mean value of the control cells. $n = 50$ cells per cell line and condition. Data in graphs are mean ± s.d., analysed by unpaired two-sided *t*-test. **f**, GSEA in cells with *TP53* mutations (Hep3B and Huh7 cells) or wild-type *TP53* (SK-Hep1 and BJ cells), treated with 10 μM XL413 for 4 days. DSB, double-strand break; FA, Fanconi anaemia; HR, homologous recombination. For gel source images, see Supplementary Fig. 1. Data in **a**, **b** are representative of three independent biological experiments. Data in **d**, **e** are representative of two independent biological experiments.
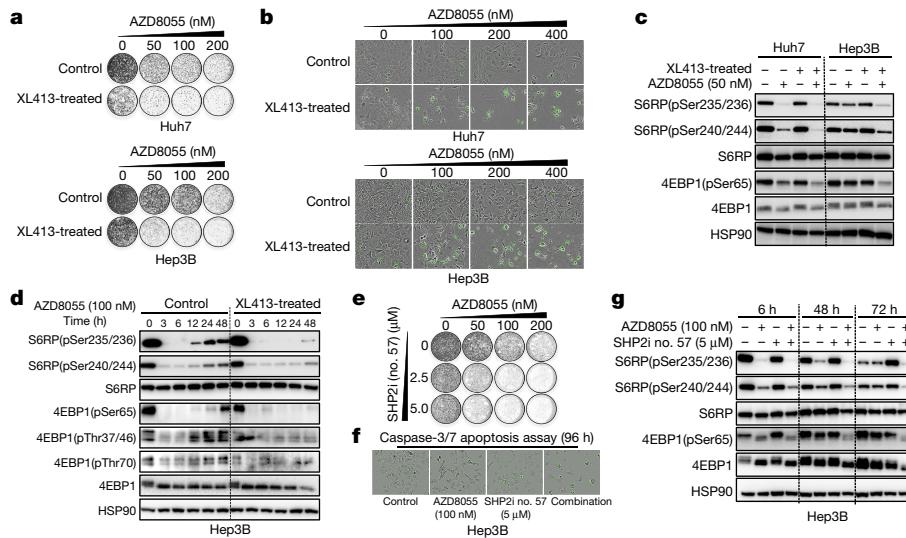
**Fig. 3 | AZD8055 selectively triggers apoptosis in XL413-induced senescent cells. a**, Hep3B and Huh7 cells were treated with 10 μM XL413 or vehicle for 10 days, before sequential exposure to increasing concentrations of AZD8055 for 5–7 days in colony-formation assays. **b**, Apoptotic cells were determined by caspase-3 and caspase-7 (caspase-3/7) apoptosis assay, 96 h after treatment with AZD8055. **c**, Control cells and XL413-induced senescent cells were treated with AZD8055 for 48 h before western blot analyses with the indicated antibodies. S6RP(pSer235/236), S6RP phosphorylated at Ser235 and Ser236; S6RP(pSer240/244), S6RP phosphorylated at Ser240 and Ser244; 4EBP1(pSer65), 4EBP1 phosphorylated at Ser65. **d**, Control cells and XL413-induced senescent cells were treated with AZD8055, and cell lysates

were collected at the indicated time points before western blot analyses with the indicated antibodies. 4EBP1(pThr37/46), 4EBP1 phosphorylated at Thr37 and Thr46; 4EBP1(pThr70), 4EBP1 phosphorylated at Thr70. **e**, **f**, Long-term colony formation assays, and caspase-3 and caspase-7 apoptosis assays, showing the synergistic effect of mTOR and SHP2 inhibitors on the proliferation of Hep3B cells. SHP2i no. 57, SHP2 inhibitor (compound no. 57; Methods). **g**, Hep3B cells were treated with AZD8055, SHP2 inhibitor or a combination of both drugs at the indicated time points, before western blot analysis with the indicated antibodies. For gel source images, see Supplementary Fig. 1. Data in **a**–**f** are representative of three independent biological experiments. Data in **g** are representative of two independent biological experiments.

translational use in the clinic (data not shown). We therefore sought to identify less-toxic compounds to selectively kill senescent liver cancer cells using a library screen of G-protein-coupled receptor (GPCR) compounds in proliferating and in XL413-treated senescent Huh7 cells (Extended Data Fig. 6a). Of these compounds, only the anti-depressant sertraline exhibited differential effects on proliferating versus XL413-induced senescent cells (Extended Data Fig. 6b, c); sertraline had modest effects on proliferating cells, but induced substantial apoptosis after treatment with XL413 (Extended Data Fig. 6d–f).

The concentration of sertraline needed to induce apoptosis of senescent cells precludes its clinical use. We therefore explored the mechanism through which sertraline selectively induces apoptosis in XL413-treated senescent cells. We analysed signalling pathways in cells treated with sertraline, and found that this treatment leads to inhibition of S6RP and 4EBP1 phosphorylation in XL413-induced senescent cells (Extended Data Fig. 6g). This suggests that the apoptotic effects of sertraline may involve regulation of mTOR signalling, as has previously been reported[18]. Consistently, gene-set enrichment analyses (GSEA) on RNA-sequencing data from cells treated sequentially with XL413 and sertraline indicated the enrichment of a gene set related to the downregulation of mTOR signalling (Extended Data Fig. 6h).

To explore whether mTOR inhibitors may be used as effective drugs in our XL413-induced senescence models, we analysed the activity of two mTOR inhibitors (AZD8055 and AZD2014). Both of these inhibitors induced apoptosis in XL413-treated liver- and lung cancer cells with *TP53* mutations, but only limiting the proliferation of untreated cells (Fig. 3a, b, Extended Data Fig. 6i–k). As expected, sequential treatment with AZD8055 did not lead to apoptosis in non-senescent liver cancer cells with wild-type *TP53* that were pre-treated with XL413 (Extended Data Fig. 6l). Importantly, mTOR signalling was further inhibited in XL413-induced senescent cells exposed to AZD8055 or AZD2014, as compared to proliferating cells (Fig. 3c, Extended Data Fig. 6m).

mTOR blockade results in a feedback-loop reactivation of mTOR signalling through the engagement of receptor tyrosine kinases, which

thus limits the efficacy of mTOR inhibitors[6]. We explored the feedback activation of mTOR signalling in time-course experiments, and found that the rapid reactivation of mTOR—as judged by the phosphorylation of S6RP and 4EBP1 at multiple sites—was observed in proliferating, but not in senescent, Hep3B cells (Fig. 3d). This feedback reactivation loop may stem from both transcriptional and biochemical activation of EGFR, PDGFRβ and IGF-1R, which leads to an increase in the phosphorylation of SHP2—this latter process is disrupted in Hep3B cells treated with XL413 (Extended Data Fig. 7a–c). Combining mTOR and SHP2 inhibitors resulted in an inhibition of the feedback reactivation of mTOR signalling and caused cell death in proliferating Hep3B cells (Fig. 3e–g), which indicates that suppression of mTOR reactivation is critical for the induction of apoptosis in senescent cells. In support of these findings, inhibition of mTOR also induced the activation of AKT in proliferating cells, and inhibition of AKT synergized with mTOR blockade to induce cell death (Extended Data Fig. 7d–f). Oncogene-induced senescent primary fibroblasts were insensitive to treatment with AZD8055 (Extended Data Fig. 8a, b), which indicates that not all senescent cells are killed by inhibition of mTOR. Importantly, feedback reactivation of mTOR was not impaired in cisplatin- or alisertib-induced senescent Hep3B cells and, consequently, no cell death was observed following the inhibition of mTOR in these cells (Extended Data Fig. 8c–e). These data indicate that the efficacy of mTOR inhibitors is dependent on context, and relies on CDC7 inhibition.

To assess whether our in vitro findings could be recapitulated in vivo, we generated Huh7 and MHCC97H xenografts. XL413-treated tumours showed increased DNA damage and SA-β-gal+ senescent cells as compared to tumours treated with vehicle, AZD8055 or a combination of XL413 and AZD8005, which indicates that the inhibition of CDC7 induces senescence in vivo (Extended Data Fig. 9a). No SA-β-gal staining was observed in SK-Hep1 tumours with wild-type *TP53* that were treated with XL413, which is consistent with the notion that the inhibition of CDC7 induces senescence only in a background of *TP53* mutation (Extended Data Fig. 9b). Compared to treatment with sorafenib, treatment with a combination of XL413 and AZD8055
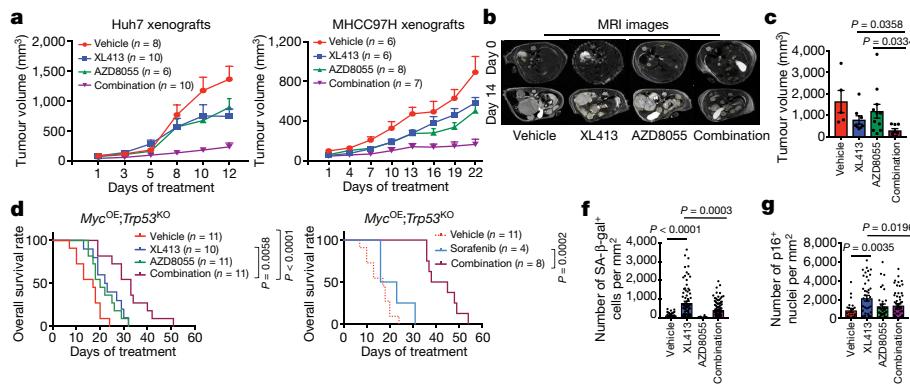
**Fig. 4 | Inhibition of CDC7 induces senescence in vivo, and suppresses tumour growth when combined with mTOR inhibition in multiple models of liver cancer. a**, Huh7 and MHCC97H cells were grown as tumour xenografts in BALB/c nude mice. Longitudinal progression of tumour volume in mice bearing Huh7 and MHCC97H tumours that were treated for 12 or 22 days, respectively, with vehicle, XL413, AZD8055 or both drugs combined. Graph shows mean ± s.e.m. **b–g**, Analyses of response to treatment with both drugs in the $Myc^{OE}$;$Trp53^{KO}$ somatic mouse model of HCC. **b**, Representative magnetic resonance imaging (MRI) images out of nine independent experimental cohorts, of day 0 and day 14 of mice enrolled in treatment with vehicle, XL413, AZD8055 or a combination of both drugs. The yellow line indicates the visible tumour area that was used to calculate the tumour volume. **c**, Tumour volumes were calculated on the basis of MRI images from mice bearing HCC with a matched initial tumour volume. Graph shows mean ± s.e.m. from

mice treated with vehicle ($n = 5$), XL413 ($n = 8$), AZD8055 ($n = 11$) or a combination of the two drugs ($n = 8$) at the intermediate time point after the initiation of treatment (days 14–16 in matched treatment groups). Unpaired two-sided $t$-test. **d**, Survival curve generated from mice bearing $Myc^{OE}$;$Trp53^{KO}$ tumours, treated with vehicle ($n = 11$; median survival 17 days), XL413 ($n = 10$; median survival of 22.5 days), AZD8055 ($n = 11$; median survival of 20 days) or a combination of the two drugs ($n = 11$; median survival of 33 days). **e**, Survival curve generated from independent cohorts of mice bearing $Myc^{OE}$;$Trp53^{KO}$ HCC, treated with sorafenib ($n = 4$; median survival of 19.5 days) or a combination of XL413 and AZD8055 ($n = 8$; median survival of 41.5 days). The vehicle group from **d** is used as a reference. **d, e**, Statistical significance was calculated using a two-sided log-rank test. **f, g**, Graphs show mean ± s.e.m. of the number of SA-β-gal+ (**f**) or p16+ (**g**) cells per tumour nodule per mm². Unpaired two-sided $t$-test. For sample sizes, see Methods.

elicited a more-effective inhibition of growth, and combination-treated xenografts with *TP53* mutations displayed diminished proliferation and phosphorylation of 4EBP1 that was associated with increased apoptosis (Fig. 4a, Extended Data Fig. 9c–g).

In immune-competent, somatic mouse models of HCC[19] (Extended Data Fig. 10a), treatment with XL413 induced senescence specifically in *Trp53*-deficient tumours (overexpression of *Myc* and knockout of *Trp53*; $Myc^{OE}$;$Trp53^{KO}$) but not in $Myc^{OE}$;$Pten^{KO}$ tumours (Extended Data Fig. 10b). Mice bearing $Myc^{OE}$;$Trp53^{KO}$ tumours that received XL413 or AZD8055 monotherapy showed a modest reduction in tumour volume and increased mouse lifespan, whereas treatment with XL413 combined with AZD8055 was well-tolerated, significantly reduced tumour burden and increased survival compared to either monotherapy or to treatment with sorafenib in this model of aggressive HCC (Fig. 4b–e, Extended Data Fig. 10c–f). Importantly, the number of SA-β-gal+ and p16 (INK4A)+ cells was decreased in the combination-treated group, which suggests that senescent cells were efficiently eliminated by treatment with AZD8055 (Fig. 4f, g, Extended Data Fig. 10g, h). An influx of macrophages (CD11b+Ly6C−Ly6G−), CD4+ T cells and increased proliferation of CD4+ and CD8+ T cells were observed after treatment with XL413 at the intermediate time-point of treatment. These changes were largely lost in the combination-treated groups, and in XL413-treated endpoint tumours (Extended Data Fig. 10i). Withdrawal of XL413 after the induction of senescence in vivo did not alter the absolute number of senescent cells, which suggests that infiltrating immune cells were unable to efficiently clear senescent cells (Extended Data Fig. 10j).

Our data indicate that pro-senescence therapy with a CDC7 inhibitor, combined with mTOR inhibitor, may deliver clinical benefit in liver cancer by alleviating both the cell-autonomous[20] and non-cell-autonomous[21] attributes of senescent cells—thus reducing the risk of tumour relapse. Although immune surveillance was mobilized, it had limited effect after inhibition of CDC7. It will be worthwhile to investigate whether combining immunotherapy (which has demonstrated activity in HCC[22]) with pro-senescence therapy can activate the cytotoxic potential of recruited immune cells in tumours that have been treated with pro-senescence therapy.

## Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-019-1607-3.

1. Bray, F. et al. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J. Clin.* **68**, 394–424 (2018).
2. Llovet, J. M. et al. Hepatocellular carcinoma. *Nat. Rev. Dis. Primers* **2**, 16018 (2016).
3. Llovet, J. M. et al. Sorafenib in advanced hepatocellular carcinoma. *N. Engl. J. Med.* **359**, 378–390 (2008).
4. Wang, L. et al. High-throughput functional genetic and compound screens identify targets for senescence induction in cancer. *Cell Reports* **21**, 773–783 (2017).
5. Sieben, C. J., Sturmlechner, I., van de Sluis, B. & van Deursen, J. M. Two-step senescence-focused cancer therapies. *Trends Cell Biol.* **28**, 723–737 (2018).
6. Rodrik-Outmezguine, V. S. et al. mTOR kinase inhibition causes feedback-dependent biphasic regulation of AKT signaling. *Cancer Discov.* **1**, 248–259 (2011).
7. Zucman-Rossi, J., Villanueva, A., Nault, J. C. & Llovet, J. M. Genetic landscape and biomarkers of hepatocellular carcinoma. *Gastroenterology* **149**, 1226–1239 (2015).
8. Chang, J. et al. Clearance of senescent cells by ABT263 rejuvenates aged hematopoietic stem cells in mice. *Nat. Med.* **22**, 78–83 (2016).
9. Ogrodnik, M. et al. Cellular senescence drives age-dependent hepatic steatosis. *Nat. Commun.* **8**, 15691 (2017).
10. Wang, C. et al. A CRISPR screen identifies CDK7 as a therapeutic target in hepatocellular carcinoma. *Cell Res.* **28**, 690–692 (2018).
11. Koltun, E. S. et al. Discovery of XL413, a potent and selective CDC7 inhibitor. *Bioorg. Med. Chem. Lett.* **22**, 3727–3731 (2012).
12. Montagnoli, A., Moll, J. & Colotta, F. Targeting cell division cycle 7 kinase: a new approach for cancer therapy. *Clin. Cancer Res.* **16**, 4503–4508 (2010).
13. Cancer Genome Atlas Research Network. Comprehensive and integrative genomic characterization of hepatocellular carcinoma. *Cell* **169**, 1327–1341 (2017).
14. Fridman, A. L. & Tainsky, M. A. Critical pathways in cellular senescence and immortalization revealed by gene expression profiling. *Oncogene* **27**, 5975–5987 (2008).
15. Faget, D. V., Ren, Q. & Stewart, S. A. Unmasking senescence: context-dependent effects of SASP in cancer. *Nat. Rev. Cancer* **19**, 439–453 (2019).

16. Dauch, D. et al. A MYC-aurora kinase A protein complex represents an actionable drug target in p53-altered liver cancer. *Nat. Med.* **22**, 744–753 (2016).
17. Eggert, T. et al. Distinct functions of senescence-associated immune responses in liver tumor surveillance and tumor progression. *Cancer Cell* **30**, 533–547 (2016).
18. Lin, C. J., Robert, F., Sukarieh, R., Michnick, S. & Pelletier, J. The antidepressant sertraline inhibits translation initiation by curtailing mammalian target of rapamycin signaling. *Cancer Res.* **70**, 3199–3208 (2010).
19. Keng, V. W. et al. A conditional transposon-based insertional mutagenesis screen for genes associated with mouse hepatocellular carcinoma. *Nat. Biotechnol.* **27**, 264–274 (2009).
20. Milanovic, M. et al. Senescence-associated reprogramming promotes cancer stemness. *Nature* **553**, 96–100 (2018).
21. Georgilis, A. et al. PTBP1-mediated alternative splicing regulates the inflammatory secretome and the pro-tumorigenic effects of senescent cells. *Cancer Cell* **34**, 85–102 (2018).
22. Flynn, M. J., Sayed, A. A., Sharma, R., Siddique, A. & Pinato, D. J. Challenges and opportunities in the clinical development of immune checkpoint inhibitors for hepatocellular carcinoma. *Hepatology* **69**, 2258–2270 (2019).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## METHODS

No statistical methods were used to predetermine sample size. The experiments were not randomized and investigators were not blinded to allocation during experiments and outcome assessment.

**Cell lines.** The human liver cancer cell lines, Hep3B, Huh7, HepG2, SNU182, SNU398, SNU449, Huh6, SK-Hep1 and PLC/PRF/5 were provided by Erasmus University. MHCC97H and HCCLM3 were provided by the Liver Cancer Institute of Zhongshan Hospital. The majority of liver cancer cell lines were established from HCC. Among these cell lines, SK-Hep1 was established from an endothelial tumour in the liver and Huh6 is a hepatoblastoma cell line. Liver cancer cells were cultured in DMEM with 10% FBS, glutamine and penicillin–streptomycin (Gibco) at 37 °C and 5% $CO_2$. The liver cancer cell lines were authenticated by applying short-tandem-repeat DNA profiling. HCT116 ($TP53^{+/+}$ and $TP53^{-/-}$) cells were provided by B. Vogelstein. hTERT immortalized BJ fibroblasts and retinal pigment epithelial cells (RPE-1) were provided by X. Qiao (Netherlands Cancer Institute). TIG-3 immortalized with hTERT and MCF-10A cells were provided by L. Li (Netherlands Cancer Institute). Two mouse cell models of liver cancer with different genetic backgrounds ($Nras^{G12V}$;$Myc^{OE}$;$Trp53^{-/-}$ and $Nras^{G12V}$;$Myc^{OE}$;$Cdkn2a^{ARF-/-}$) were provided by L. Zender (University Hospital Tubingen). Mycoplasma contamination was excluded using a PCR-based method.

**Compounds and antibodies.** XL413 (S7547), BMS265246 (S2014), ON-01910 (S1362), PD0166285 (S8148), LDC000067 (S7461), PF-03814735 (S2725), D 4476 (S7642), VE-821 (S8007), AZD8055 (S1555), AZD2014 (S2783), AZD6738 (S7693) and MK-8776 (2735) were purchased from Selleck Chemicals. THZ531 (A8736) was purchased from ApexBio. XL413 (205768), BLU9931 (206192) and LY3177833 (206762) were purchased from MedKoo. TAK-931 (CT-TAK931) was purchased from Chemietek. XL413 (A13677) was also purchased from AdooQ BIOSCIENCE. The SHP2 inhibitor used in this study is covered by a patent application (WO 2015/107495A1; compound no. 57) and was synthesized as previously described[23].

Antibodies against HSP90 (sc-7947, sc-13119), p53 (sc-126), p21 (sc-6246), and SHP2 (sc-280) were purchased from Santa Cruz Biotechnology. Antibodies against CDC7 (ab77668), p-MCM2 (ab109133, ab133243), MCM2 (ab4461), p-SHP2 (ab62322), PCNA (ab2426), and cleaved caspase-3 (ab2303) were purchased from Abcam. Antibodies against γH2AX (no. 9718), p-S6RP (no. 4856, no. 5364), S6RP (no. 2317), p-4EBP1 (no. 9456, no. 2855, no. 9455), 4EBP1 (no. 9644), p-IGF-1R/INSR (no. 3024), IGF-1R (no. 9750), p-PDGFRβ (no. 3161), PDGFRβ (no. 4564), p-AKT (no. 4060) and AKT (no. 2920) were purchased from Cell Signalling. EGFR antibody (610017) was purchased from BD Biosciences. H3K9me3 antibody (49-1008) and p-EGFR (44-788) were from Thermo Fisher Scientific.

**Pooled 'stress lethal' CRISPR screen.** For the design of the kinome CRISPR library, 5,971 gRNAs targeting 504 human kinases, 10 essential genes and 50 non-targeting gRNAs were selected. Oligonucleotides with gRNA sequences flanked by adaptors were ordered from CustomArray, and cloned as a pool by GIBSON assembly in LentiCRISPRv2.1. The kinome CRISPR library was introduced to Hep3B and Huh7 cells by lentiviral transduction. Cells stably expressing gRNA were cultured for 14 days. The abundance of each gRNA in the pooled samples was determined by Illumina deep-sequencing. gRNAs prioritized for further analysis were selected by the fold depletion of abundance in the day-14 sample compared with that in the day-0 sample, using previously described methods[24].

**Compound screens.** *Induction of senescence screen.* We performed a compound screen including 10 small-molecule inhibitors that targeting the 14 hits identified in the CRISPR screen. The compounds used for this screen are described in Fig. 1c. Each compound was evaluated in two liver cancer cell lines (Hep3B and Huh7) and two non-transformed cell lines (BJ and RPE-1) using five different concentrations. The screens were performed in three replicates of each cell line. SA-β-gal staining was performed after 4 days of treatment.

*Killing senescent-cell screen.* Cells were screened for sensitivity against a panel of 260 small-molecule inhibitors from a GPCR compound library (L2200, Selleck Chemicals). In brief, Huh7 cells were treated with 10 μM XL413 for 5 days, and then control cells and XL413-treated cells were plated in 96-well plates. All compounds from GPCR library were tested at four concentrations. Each plate included 8 wells containing DMSO (as a negative control) and 8 wells containing 10 μM PAO (as a positive control). The cell viability in each well was determined using CellTiter-Blue reagent (Promega). The relative survival of control cells and XL413-treated senescent cells in the presence of drug was normalized against control conditions (untreated cells) after subtraction of background signal.

**SA-β-gal staining.** SA-β-gal staining was performed either in 6-well or 96-well plates (for in vitro studies), on 10-μm-thick cryosections from xenografted tumours or on 8-μm-thick cryosections from hydrodynamic-tail-vein-injection (HDTVi)-generated $Myc^{OE}$;$Trp53^{KO}$ tumours, using a commercial kit (Sigma) following the manufacturer's instructions.

**Protein lysate preparation and western blots.** Cells were washed with PBS and lysed with RIPA buffer supplemented with complete protease inhibitor (Roche) and phosphatase inhibitor cocktails II and III (Sigma). Protein quantification was performed with the BCA protein assay kit (Pierce). All lysates were freshly prepared and processed with Novex NuPAGE gel electrophoresis systems (Thermo Fisher Scientific), followed by western blotting.

**Immunohistochemical staining.** Specimens of HCC were obtained from 80 patients (from 26 to 76 years old) who underwent curative surgery in Eastern Hepatobiliary Hospital of the Second Military Medical University. Patients were not subject to any preoperative anti-cancer treatment. Ethical approval was obtained from the Eastern Hepatobiliary Hospital Research Ethics Committee, and written informed consent was obtained from each patient. Of these cases, 12 patients are female and 68 patients are male. Fifty-nine patients had a background of HBV infection. Clinical information—including tumour number, diameter of tumour, tumour differentiation, serum AFP, status of cancer recurrence, disease-free survival and death from recurrence—was collected. For immunohistochemical analysis, formalin-fixed paraffin-embedded samples from patients with HCC were probed with CDC7 antibody (ab77668, Abcam). Formalin-fixed paraffin-embedded samples were also obtained from xenograft tumours or tumours from immunocompetent somatic mouse models, and then probed with antibodies against PCNA (ab2426, Abcam), cleaved caspase-3 (ab2303, Abcam), p-4EBP1 (no. 2855, Cell Signaling) or p16 (ab54210, Abcam). Following incubation with the primary antibodies, positive cells were visualized using DAB+ as a chromogen. For the analysis of p16 and SA-β-gal staining, slides were digitally processed using the Aperio ScanScope (Aperio) at a magnification of 20×. Nodule size was drawn by hand in HALO image-analysis software (Indica Labs) and an algorithm was designed with the Multiplex IHC v.1.2 module to quantify the number of positive cells[25] either as absolute or per $mm^2$ (as indicated in figure legends).

**Long-term cell-proliferation assays (colony formation).** Cells were cultured and seeded into 6-well plates at a density of $5 \times 10^3$ to $4 \times 10^4$ cells per well, depending on growth rate, and were cultured in medium containing the indicated drugs for 10–14 days. The medium was changed twice a week. Cells were fixed with 4% formaldehyde in PBS and stained with 0.1% crystal violet diluted in water.

**Plasmids.** All lentiviral shRNA vectors were retrieved from the arrayed The RNAi Consortium human genome-wide shRNA collection. These shRNAs were as follows: *CDC7* shRNA no. 1: TRCN0000003168_ CCGGGCCA CAGCACAGTTACAAGTACTCGAGTACTTGTAACTGTGCTG TGGCTTTTT; *CDC7* shRNA no. 2: TRCN0000196542_CCGGGAAGCTTTGTTGCAT CCATTTCT CGAGAAATGGATGCAACAAAGCTTCTTTTTTG; *TP53* shRNA no. 1: TRCN0000010814_CCGGGAGGGATGTTTGGGA GATGTACTCGAGTACA TCTCCCAAACATCCCTCTTTTT; *TP53* shRNA no. 2: TRCN0000003754_ CCGGTCAGACCTATGGAAACTACTTCTCGAGAAGT AGTTTCCATAGGTCTGATTTTT; and *TP53* shRNA no. 3: TRCN0000003755_ CCGGGTCCAGATGAAGCTCCCAGAACTC GAGTTCTGGGAGCTTCATCT GGACTTTTT

**Incucyte cell-proliferation assay and apoptosis assay.** Indicated cell lines were seeded into 96-well plates at a density of 1,000–8,000 cells per well, depending on growth rate and the design of the experiment. About 24 h later, drugs were added at the indicated concentrations using the HP D300 Digital Dispenser (HP). Cells were imaged every 4 h using the Incucyte ZOOM (Essen Bioscience). Phase-contrast images were analysed to detect cell proliferation on the basis of cell confluence. For cell apoptosis, caspase-3 and caspase-7 green apoptosis-assay reagent was added to the culture medium, and cell apoptosis was analysed on the basis of green fluorescent staining of apoptotic cells.

**RNA sequencing.** RNA (one sample per cell line per condition) was isolated using Trizol, and cDNA libraries were sequenced on an Illumina HiSeq2500 to obtain 65-bp single-end sequence reads. Reads were aligned to the GRCh38 human reference genome. GSEA was performed using GSEA software as previously described[26]. The 'FRIDMAN_SENESCENCE_UP'[14] gene set was used to assess the enrichment of senescence-associated genes in XL413-treated versus control cells. Gene sets related to DNA-damage repair were used to assess the enrichment of genes associated with DNA-damage repair in the XL413-treated versus control cells. Enrichment scores were corrected for gene-set size (normalized enrichment score). The 'PENG_RAPAMYCIN_RESPONSE_DN'[27] gene set was used to assess the enrichment of downregulation of mTOR signalling in liver cancer cells that had been sequentially treated with XL413 and sertraline, versus control cells. The P value estimates the statistical significance of the enrichment score for a single gene set, as previously described[26]. Exact P values are shown in the figures, unless the P value < 0.001.

**Immunofluorescence and image analysis.** For immunofluorescence microscopy, cells were seeded on glass coverslips and cultured in the presence of 10 μM XL413 for 7 days. Cells were fixed in 2% paraformaldehyde and permeabilized with 0.2% Triton X-100 for 5 min, blocked with PBS containing 2% bovine serum albumin (Sigma-Aldrich) for 45 min and subsequently incubated with H3K9me3 antibody (Thermo Fisher Scientific, 49-1008) and goat anti-rabbit Alexa Fluor 488 (Invitrogen; 1:200) for 1 h, respectively. Nuclei were stained with 4,6-diamidino-2-phenylindole. Samples were mounted on glass slides in Mowiol after three

washing steps with PBS. Images were acquired with a Leica TCS SP5 confocal microscope with a 63× (NA 1.4) oil objective. Image processing was performed using ImageJ software.

**Neutral comet assay.** To detect DNA double-strand breaks, neutral comet assays were performed as previously described[28]. In bief, cells were collected and embedded in 1% low-gelling-temperature agarose (Sigma-Aldrich). A cell suspension was used to make gels onto comet assay slides (Trevigen). Cells in the agarose gels were lysed at 37 °C in lysis buffer (2% sarkosyl, 0.5M $Na_2EDTA$ and 0.5 mg/ml proteinase K) overnight. Subsequently, slides were washed 3 times for 30 min at room temperature in electrophoresis buffer (90 mM Tris-HCl pH 8.5, 90 mM boric acid and 2 mM $Na_2EDTA$). Electrophoresis was performed for 25 min at 20 V in electrophoresis buffer. Afterwards, slides were washed once with MQ, and DNA was stained using 2.5 µg/ml propidium iodide in MQ. Individual comets were imaged with a Zeiss AxioObserver Z1 inverted microscope. Tail moments of individual comets were assessed using the CASP software. For each condition, at least 50 cells were analysed.

**Time-lapse live imaging.** To allow visualization of chromosomes, cells were transduced with a histone H2B–GFP (LV-GFP, Addgene plasmid no. 25999). Cells were then plated 24 h before starting the microscope acquisition. XL413 (10 µM) was added in the medium 1 h before starting the movie. Cells were filmed over 96 h and images were taken every 10 min. For each condition filmed, five different fields were selected. In each field, we randomly choose and followed cells entering in mitosis. Nuclear envelope breakdown was used as an indicator of the onset of mitotic division.

**Quantitative reverse-transcription PCR.** Total RNA was extracted from cells using Trizol reagent from Invitrogen or Quick-RNA MiniPrep from Zymo Research. cDNA synthesis was performed using Maxima Universal First Strand cDNA Synthesis Kit from Thermo Scientific. Quantitative PCR reactions were performed with FastStart Universal SYBR Green Master (Rox) from Roche. The experiments were performed according to the manufacturer's instructions. The sequences of the primers used for quantitative reverse-transcription PCR (RT–qPCR) analyses were as follows: *IL6* forward, ACTCACCTCTTCAGAACGAATTG; *IL6* reverse, CCATCTTTGGAAGGTTCAGGTTG; *IL8* forward, TTTTGC CAAGGAGTGCTAAAGA; *IL8* reverse, AACCCTCTGCACCCAGTTTTC; *MMP1* forward, TTGTGGCCAGAAAACAGAAA; *MMP1* reverse, TTCGGGGAGAA GTGATGTTC; *MMP3* forward, CAATTTCATGAGCAGCAACG; *MMP3* reverse, AGGGATTAATGGAGATGCCC; *CXCL1* forward, CTTCCTCCT CCCTTCTGGTC; *CXCL1* reverse, GAAAGCTTGCCTCAATCCTG; *CXCL10* forward, GCTGATGCAGGTACAGCGT; *CXCL10* reverse, CACCATGAA TCAAACTGCGA; *EGFR* forward, AGGCACGAGTAACAAGCTCAC; *EGFR* reverse, ATGAGGACATAACCAGCCACC; *IGF1R* forward, TCGACATCCGCAACGACTATC; *IGF1R* reverse, CCAGGGCGTA GTTGTAGAAGAG; *INSR* forward, AAAACGAGGCCCGAAGATTTC; *INSR* reverse, GAGCCCATAGACCCGGAAG; *PDGFRB* forward, AGC ACCTTCGTTCTGACCTG; *PDGFRB* reverse, TATTCTCCCGTGTCTAGCCCA; *GAPDH* forward, AAGGTGAAGGTCGGAGTCAA; *GAPDH* reverse, AATGAAGGGGTCATTGATGG. All reactions were run in triplicate.

**Human phospho-receptor tyrosine kinase array.** Phospho-receptor tyrosine kinase (RTK) arrays were used to analyse alterations of kinase signalling in response to treatment with AZD8055 in Hep3B cells, according to the manufacturer's instructions (R&D systems).

**Xenografts.** All mice were manipulated according to protocols approved by the Shanghai Medical Experimental Animal Care Commission and Shanghai Cancer Institute. The maximum permitted tumour volume was 2,000 mm³. Huh7 and MHCC97H cells (5 × 10⁶ cells per mouse) were injected subcutaneously into the right posterior flanks of 6-week-old BALB/c nude mice (male, 6–10 mice per group). Tumour volume, based on calliper measurements, was calculated by the modified ellipsoidal formula: tumour volume =1/2(length × width²). After tumour establishment, mice were randomly assigned to 6 days per week treatment with vehicle, XL413 (50–100 mg/kg, oral gavage), AZD8055 (10–20 mg/kg, oral gavage) or a combination in which each compound was administered at the same dose and schedule as the single agent. For sorafenib-treatment assay, Huh7 and MHCC97H-pLKO cells (5 × 10⁶ cells per mouse) were injected subcutaneously into the right posterior flanks of 6-week-old BALB/c nude mice (male, 6 per group). Mice were randomly assigned to treatment 6 days per week with vehicle or sorafenib (30 mg/kg, daily gavage). The investigators were not blinded to allocation during experiments and outcome assessment.

**Immunocompetent mouse models of HCC.** All mouse study protocols were approved by the NKI Animal Welfare Body. Vectors for HDTVi were prepared using the EndoFree-Maxi Kit (Qiagen) and resuspended in a sterile 0.9% NaCl solution/plasmid mix containing 5 µg of pT3-MYC (Addgene 92046), 5 µg of pX330-p53 (Addgene 59910) or pX330-PTEN (Addgene 59909), and 2.5 µg of CMV-SB13 transposase. A total volume mix corresponding to 10% of body weight was injected via lateral tail vein in 5–7 s into 6–8-week-old female C57Bl/6 mice

(Janvier laboratories). Mice were monitored by weekly MRI after HDTVi. MRI was performed in ParaVision 6.0.1 on a 7T Bruker BioSpec 70/20 USR with a $^1$H transmit–receive volume coil. T2-weighted images were acquired under 1–2% isoflurane in air and oxygen flow using a respiratory-gated sequence with TR/TE = 2,500/25 ms, 32 × 24-mm field of view (320 × 240 matrix, resolution of 0.1 mm), 30 × 0.7-mm axial slices and 4 averages. MRI images were analysed with MIPAV ('Medical, Image, Processing, Analysis and Visualization' software) to calculate tumour volume. The investigators were not blinded to allocation during experiments and outcome assessment.

When HCCs were first visible by MRI (14–21 days after HDTVi), tumour-size-matched mice were randomized over the treatment groups: vehicle, XL413, AZD8055, a combination of XL413 and AZD8055, or sorafenib. Mice were dosed 6 days per week with vehicle, XL413 (100 mg/kg, oral gavage), AZD8055 (20 mg/kg, oral gavage), a combination in which XL413 and AZD8055 were administered at the same dose as the single agent, or with sorafenib (30mg/kg, oral gavage). For time-point analysis, mice were killed 14–16 days after the initiation of treatment, and for survival-curve and endpoint analysis, the treatment continued until mice were symptomatic (the tumour reached a total volume of ≥2 cm³).

No toxicity was observed over the monotherapy groups. Seventeen per cent of mice showed therapy-induced adverse events in the XL413 + AZD8055 treatment group and 83% of mice showed a well-tolerated response to treatment.

For quantification of SA-β-gal staining, the sample size was as follows: vehicle-treated, $n = 41$ biologically independent nodules from 7 mice; XL413-treated, $n = 81$ biologically independent nodules from 11 mice; AZD8055-treated, $n = 26$ nodules from 3 mice; and combination-treated, $n = 101$ nodules from 13 mice. For quantification of p16 staining, the sample size was as follows: vehicle-treated, $n = 23$ biologically independent nodules from 3 mice; XL413-treated, $n = 43$ biologically independent nodules from 5 mice; AZD8055-treated, $n = 37$ nodules from 3 mice; and combination-treated, $n = 59$ nodules from 8 mice.

**Flow cytometry.** Mouse livers were perfused with PBS and then dissociated into single-cell suspension using the Liver Dissociation kit (Miltenyi Biotec) and the gentleMACS Octo Dissociator, following the manufacturer's instructions. The cell suspension was passed through a 100-µm cell strainer (Corning) and then centrifuged at 300g for 10 min at 4 °C and washed 3 times in FACS buffer. Samples were incubated with anti-CD16/CD32 antibody (BD Biosciences) for 15 min and then stained with the indicated antibodies (Supplementary Table 2) following standard procedures. Samples were fixed with eBioscience fixation and permeabilization kit (Invitrogen) and Ki67 antibody was used for intracellular staining. The signal was detected by using a four-laser Fortessa flow cytometer (Becton Dickinson). Analyses were carried out using FlowJo software. The gating strategy is provided in Supplementary Information. For macrophages, CD4 and CD8 T cells, the sample size was as follows: vehicle-treated, intermediate time point $n = 7$ mice and endpoint $n = 8$ mice; XL413-treated, intermediate time point $n = 3$ mice and endpoint $n = 6$ mice; AZD8055-treated, intermediate time point $n = 3$ mice and endpoint $n = 5$; and combination-treated, intermediate time point $n = 6$ mice and endpoint $n = 7$ mice. For assessment of Ki67$^+$ cells in CD4 and CD8 T cell populations, samples sizes were as follows: vehicle-treated, intermediate time point $n = 7$ mice and endpoint $n = 8$ mice; XL413-treated, intermediate time point $n = 3$ mice and endpoint $n = 4$ mice; AZD8055-treated, intermediate time point $n = 3$ mice and endpoint $n = 5$; and combination-treated, intermediate time point $n = 6$ mice and endpoint $n = 3$ mice.

**Reporting summary.** Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

## Data Availability

Raw and processed data from the next-generation RNA sequencing of samples have been deposited in the NCBI Gene Expression Omnibus (GEO) under accession number GSE121276 and GSE121277. All other data can be found in the Source Data, Supplementary Information or are available from the corresponding authors upon reasonable request.

23. Garcia Fortanet, J. et al. Allosteric inhibition of SHP2: identification of a potent, selective, and orally efficacious phosphatase inhibitor. *J. Med. Chem.* **59**, 7773–7782 (2016).
24. Evers, B. et al. CRISPR knockout screening outperforms shRNA and CRISPRi in identifying essential genes. *Nat. Biotechnol.* **34**, 631–633 (2016).
25. Koelzer, V. H. et al. Digital image analysis improves precision of PD-L1 scoring in cutaneous melanoma. *Histopathology* **73**, 397–406 (2018).
26. Subramanian, A. et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl Acad. Sci. USA* **102**, 15545–15550 (2005).
27. Peng, T., Golub, T. R. & Sabatini, D. M. The immunosuppressant rapamycin mimics a starvation-like signal distinct from amino acid and glucose deprivation. *Mol. Cell. Biol.* **22**, 5575–5584 (2002).
28. Olive, P. L. & Banáth, J. P. The comet assay: a method to measure DNA damage in individual cells. *Nat. Protoc.* **1**, 23–29 (2006).

**Extended Data Fig. 1 | Upregulation of *CDC7* mRNA correlates with poor prognosis of patients with HCC, and *TP53* knockdown sensitizes liver cancer cells with wild-type *TP53* to the CDC7 inhibitor. a**, Thirty-eight common hits (among the top-50 most-strongly depleted hits in each cell line) were identified by CRISPR screen in Hep3B and Huh7 cells. Hits in red represent factors that are targetable with small-molecule compounds. Blue represents non-targetable hits. **b**, Western blot analysis of levels of CDC7, MCM2 and phosphorylated MCM2 in non-transformed cell lines and liver cancer cell lines. HSP90 served as a loading control. **c**, Immunohistochemical analysis showing increased expression of CDC7 in HCC tissues, compared to paired adjacent non-tumour tissues. **d**, According to the level of *CDC7* mRNA obtained from the TCGA database (*n* = 365 patients), patients with HCC were classified into 3 groups: the top 33.3% were considered as high-expression, the medium 33.3% were considered as intermediate-expression and the lowest 33.3% were considered as low-expression. Kaplan–Meier curves depicting that upregulation of *CDC7* mRNA correlates with poor prognosis of patients with HCC. Statistical significance was calculated using a two-sided

log-rank test. **e**, Liver cancer cell lines with wild-type *TP53* (SK-Hep1 and Huh6) were stably transduced with control pLKO vector or with one of three independent shRNAs that target *TP53* (labelled here as shp53 #1, #2 and #3). On the basis of knockdown efficiency, *TP53* shRNA no. 1 and *TP53* shRNA no. 3 were selected for further experiments. **f**, SK-Hep1 and Huh6 cells that express a control shRNA (pLKO) or knockdown of *TP53* (shp53) were exposed to the indicated concentrations of XL413 in colony-formation assays. Cells were fixed, stained and photographed after 10–14 days of culture. **g**, SK-Hep1 and Huh6 cells expressing control shRNA or shRNA against *TP53* were exposed to the indicated concentrations of XL413 for five days. CellTiter-Blue viability assays revealed that *TP53* knockdown synergizes with treatment with XL413 in SK-Hep1 and Huh6 cells. Graphs represent mean ± s.d. from six technical replicates. For gel source images, see Supplementary Fig. 1. Data in **a**, **b**, **e**–**g** are representative of three independent biological experiments. Data in **c** are representative images from immunohistochemical analyses using a tissue microarray containing 80 specimens of HCC.
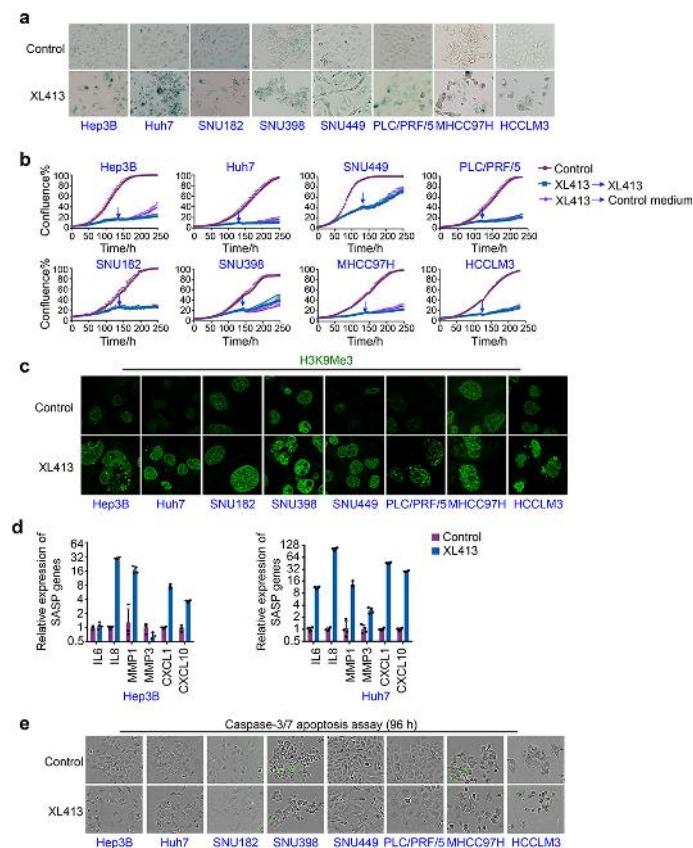
**Extended Data Fig. 2 | Inhibition of CDC7 induces senescence selectively in liver cancer cells with *TP53* mutation. a**, Liver cancer cell lines with *TP53* mutation were cultured in the presence of 10 μM XL413 for 4 days, which induces senescence (as detected by SA-β-gal staining). **b**, Growth curves (measured by Incucyte live-cell analyses) of liver cancer cell lines with *TP53* mutations that were either untreated, continuously treated with XL413 or treated with 10 μM of XL413 for 5 or 6 days before withdrawal of treatment. Graphs represent mean ± s.d. from five technical replicates. **c**, Representative images of H3K9me3 staining in liver cancer cell lines with *TP53* mutations, exposed to 10 μM XL413 for 7 days.

**d**, Treatment with XL413 induces a senescence-associated secretory phenotype (SASP) in Hep3B and Huh7 cells treated with 10 μM XL413 for 7 days. mRNA expression of genes associated with the senescence-associated secretory phenotype was determined by qRT–PCR analysis. Graphs represent mean ± s.d. from four technical replicates. **e**, Liver cancer cells were cultured in the presence of 10 μM XL413 for 4 days, and apoptotic cells were visualized by caspase-3 and caspase-7 apoptosis assay. Data in **a** are representative of three independent biological experiments. Data in **b**–**e** are representative of two independent biological experiments.
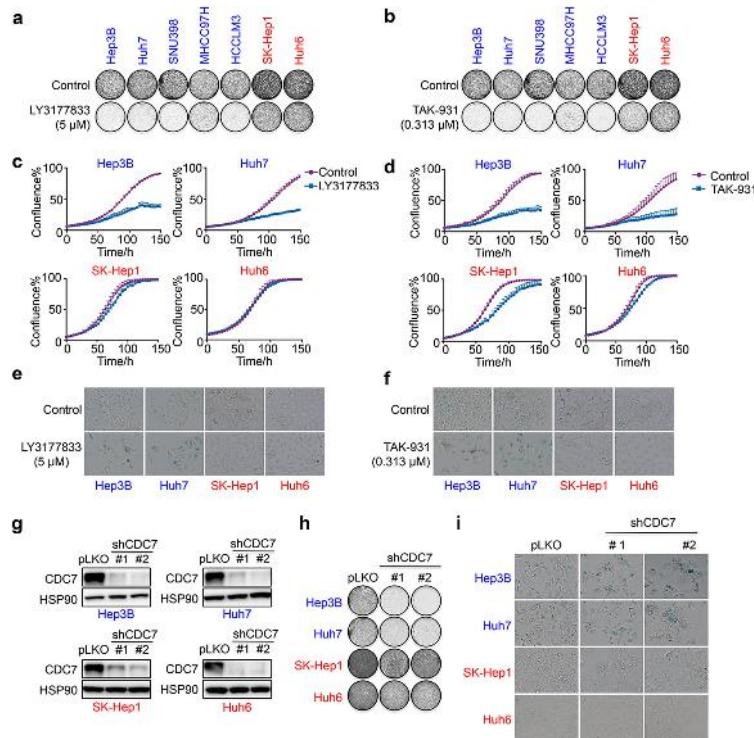
**Extended Data Fig. 3 | Pharmacological or genetic inhibition of CDC7 induces a senescent phenotype in liver cancer cells with *TP53* mutations. a, b**, Liver cancer cell lines with *TP53* mutations (Hep3B, Huh7, SNU398, MHCC97H and HCCLM3) (blue) and liver cancer cell lines with wild-type *TP53* (SK-Hep1 and Huh6) (red) were seeded at low confluence and grown in the absence or presence of the CDC7 inhibitors LY3177833 or TAK-931 at the indicated concentrations, in long-term colony-formation assays. Cells were fixed, stained and photographed after 10–14 days of culture. **c, d**, Growth curves (measured by Incucyte live-cell analyses) of liver cancer cell lines with *TP53* mutations (Hep3B and Huh7) (blue) and liver cancer cell lines with wild-type *TP53* (SK-Hep1 and Huh6) (red) exposed to LY3177833 or TAK-931. Graphs represent mean ± s.d. from four technical replicates. **e, f**, Liver cancer cells were cultured in the presence of the CDC7 inhibitors LY3177833 or TAK-931 at the indicated concentration for 4 days. SA-β-gal staining revealed that CDC7 inhibitors (LY3177833 or TAK-931) selectively induced senescence in liver cancer cells with *TP53* mutations (blue) and not in liver cancer cells with wild-type *TP53* (red). **g**, Liver cancer cell lines with *TP53* mutations (Hep3B and Huh7) and liver cancer cell lines with wild-type *TP53* (SK-Hep1 and Huh6) were stably transduced with control pLKO vector or with two independent shRNAs that target *CDC7* (labelled here shCDC7 #1 and #2) and the efficiency of CDC7 knockdown in liver cancer cell lines was evaluated by western blot. **h**, Colony-formation assays of liver cancer cell lines with *TP53* mutation (blue) and liver cancer-cell lines with wild-type *TP53* (red), with and without CDC7 knockdown, were performed. Cells were fixed, stained and photographed after ten days of culture. **i**, CDC7 knockdown induced senescence in Hep3B and Huh7 cells with *TP53* mutations, but not in SK-Hep1 and Huh6 cells, which have wild-type *TP53*. Senescence was detected by SA-β-gal staining. For gel source images, see Supplementary Fig. 1. Data in **a**–**i** are representative of three independent biological experiments.
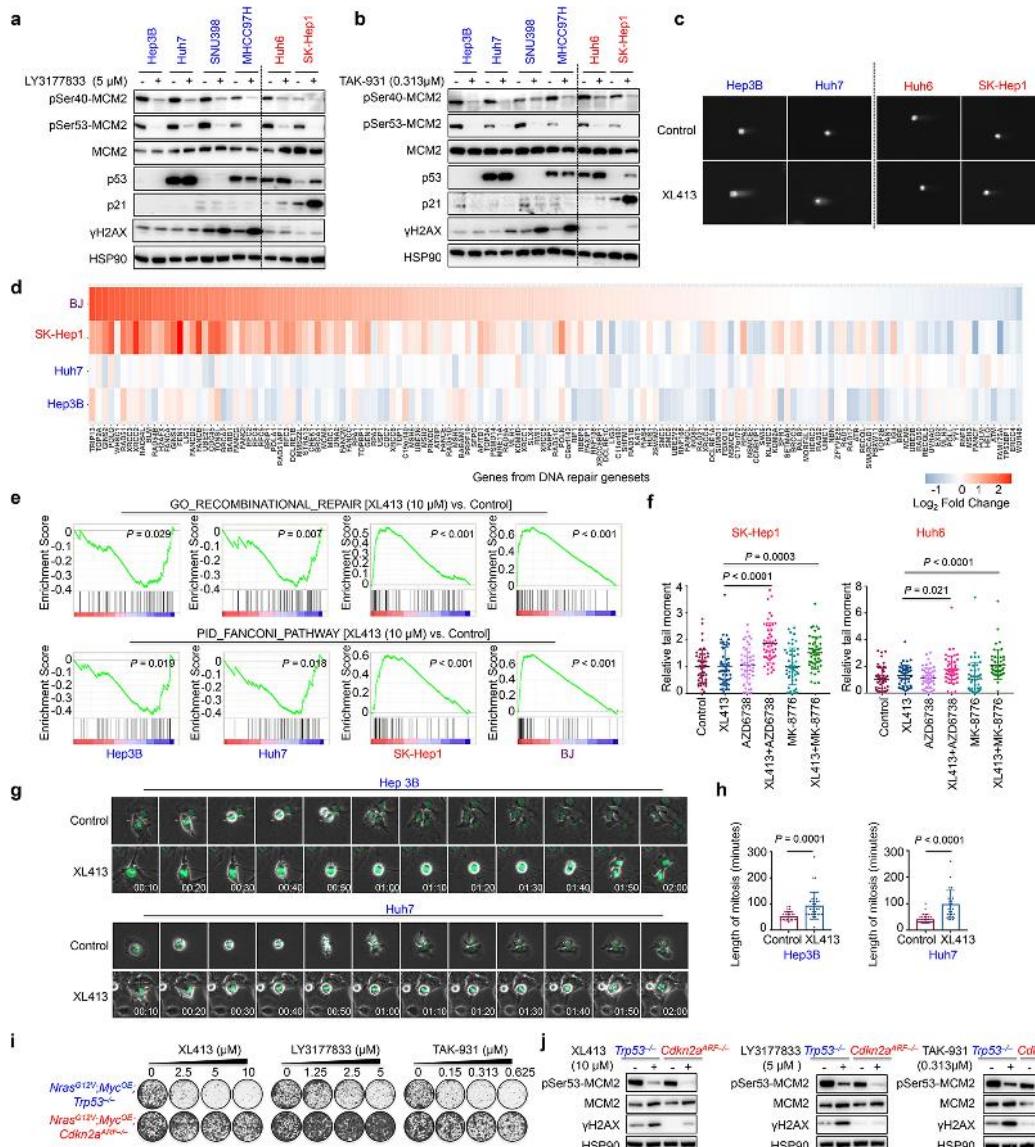
**Extended Data Fig. 4 | Inhibition of CDC7 leads to the accumulation of DNA damage specifically in liver cancer cells with *TP53* mutations.** **a**, **b**, Western blot analysis of liver cancer cell lines treated with CDC7 inhibitors (LY3177833 or TAK-931) for seven days. Inhibition of CDC7 induces the expression of the DNA-damage marker γH2AX in liver cancer cells with *TP53* mutations while lower γH2AX together with functional upregulation of p53 and p21[cip1] were observed in *TP53* wild-type liver cancer cell lines post-XL413 treatment. **c**, Representative neutral-comet-assay images of liver cancer cells with *TP53* mutations (Hep3B and Huh7) and liver cancer cells with wild-type *TP53* (SK-Hep1 and Huh6) treated with XL413 for seven days. **d**, Heat map displays fold gene-expression changes (expressed in log₂) in cells with wild-type *TP53* (BJ and SK-Hep1) and liver cancer cells with *TP53* mutations (Huh7 and Hep3B) upon treatment with XL413 (10 μM, 4 days). **e**, GSEA was performed on RNA-sequencing data from Hep3B, Huh7, SK-Hep1 and BJ cells treated with 10 μM XL413 for 4 days; this identified DNA-repair signatures (recombinational repair and Fanconi anaemia pathway) to be significantly different between cells with *TP53* mutations and cells with wild-type *TP53* (Methods). **f**, Neutral comet assays were performed on SK-Hep1

and Huh6 cells treated with 20 μM XL413 combined with AZD6738 (ATR inhibitor, 2.5 μM) or MK-8776 (CHK1 inhibitor, 2.5 μM) for 3 days. The value of tail moments in each treatment group were normalized on the basis of the mean value of the control cells (*n* = 50 cellsper cell line and condition). Graphs represent mean ± s.d., analysed with unpaired two-sided Student's *t*-test. **g**, **h**, H2B–GFP Hep3B and Huh7 cells were cultured in absence or presence of XL413 (10 μM), and time-lapse microscopy was performed over 96 h to measure the length of mitosis. Graphs represent mean ± s.d. *n* = 30 cells per cell line and condition, analysed with unpaired two-sided *t*-test. **i**, Mouse cell models of liver cancer with different genetic backgrounds (*Nras*[G12V];*Myc*[OE];*Trp53*[−/−] and *Nras*[G12V];*Myc*[OE];*Cdkn2a*[ARF−/−]) were exposed to the indicated concentrations of CDC7 inhibitors (XL413, LY3177833 or TAK-931) for seven days in colony-formation assays. **j**, Western blot analysis of mouse cell models of liver cancer treated with XL413, LY3177833 or TAK-931 for seven days. For gel source images, see Supplementary Fig. 1. Data in **a**–**c**, **f**, **g**, **h** are representative of two independent biological experiments. Data in **i**, **j** are representative of three independent biological experiments.
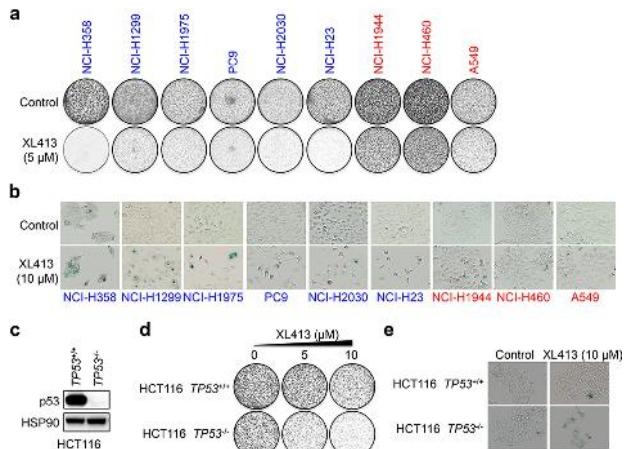
**Extended Data Fig. 5 | Inhibition of CDC7 induces senescence selectively in cancer cells with *TP53* mutations. a**, Lung-cancer cell lines with *TP53* mutations (blue) and lung-cancer cell lines with wild-type *TP53* (red) were seeded at low confluence and grown in the absence or presence of XL413 at the indicated concentration for 10–14 days in colony-formation assays. **b**, Lung-cancer cells were exposed to 10 μM XL413 for 4 days, which induces senescence selectively in cells with *TP53* mutations (as detected by SA-β-gal staining). **c**, Expression of p53 was assessed in isogenic *TP53*^−/− and *TP53*^+/+ HCT116 colon-cancer cell lines by western blot. **d**, HCT116 *TP53*^+/+ and HCT116 *TP53*^−/− cells were seeded at low confluence and grown in the absence or presence of XL413 at the indicated concentration for seven days in a colony-formation assay, to assess their proliferation capacity. **e**, HCT116 *TP53*^+/+ and HCT116 *TP53*^−/− cells were cultured in the presence of 10 μM XL413 for 4 days, and senescence was selectively induced in *TP53*^−/− HCT116 cells (as detected by SA-β-gal staining). For gel source images, see Supplementary Fig. 1. Data in **a–e** are representative of two independent biological experiments.
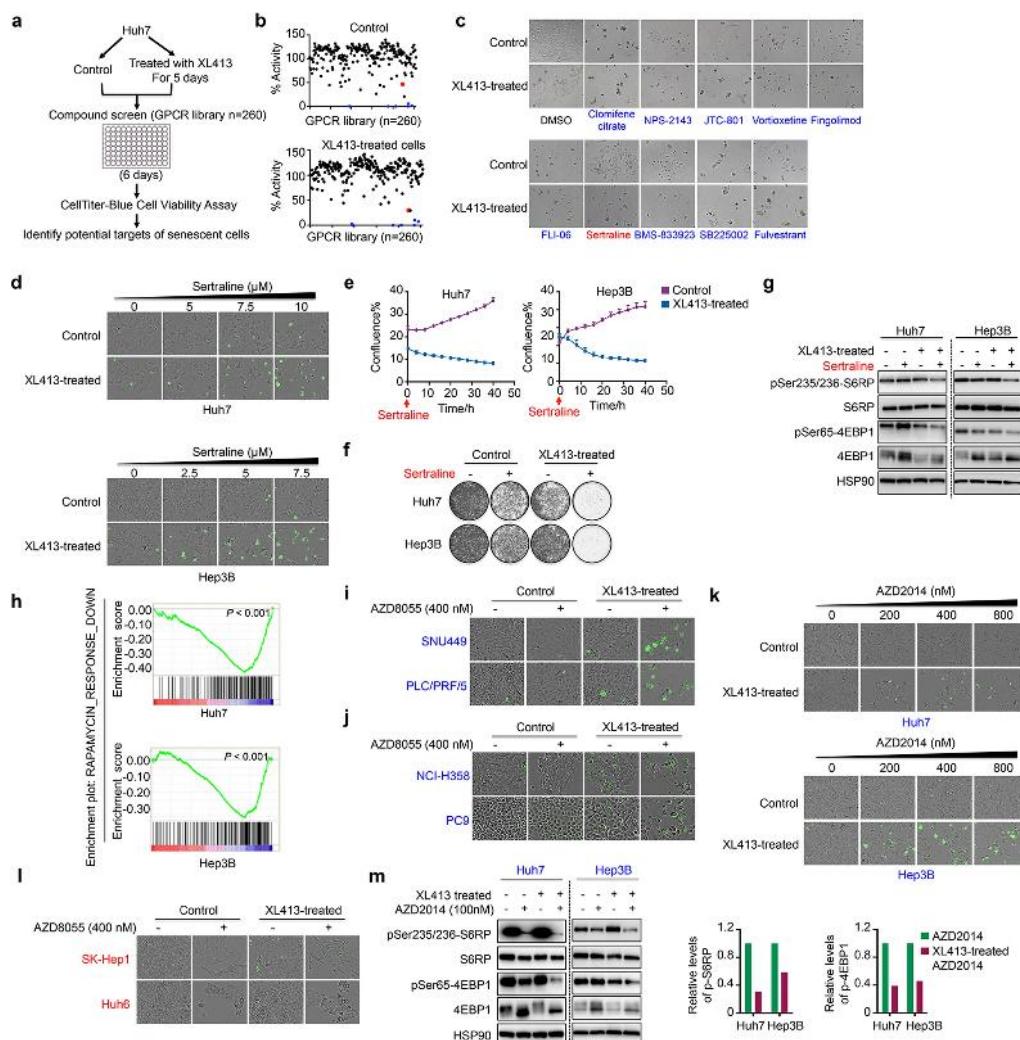
**Extended Data Fig. 6 | Sertraline selectively induces apoptosis in XL413-induced senescent cells through the suppression of mTOR signalling. a**, Schematic of the GPCR-compound screen. Huh7 cells were treated with 10 μM XL413 for 5 days before seeding in 96-well plates. All compounds were tested at four concentrations for six days, and cell viability was measured using CellTiter-Blue assay. **b, c**, Graph depicting the effects of compounds on cell viability. Each point represents a single compound, with per cent activity calculated by dividing the cell viability score in the presence of 5 μM of that compound by the mean viability of the negative control. Blue dots indicate compounds that induce cell death in both control and XL413-induced senescent cells. Sertraline (red dot) induced selective cell death in XL413-induced senescent cells. Representative images of the effects of compounds on XL413-treated and untreated cells are shown. **d**, Control cells and XL413-induced senescent cells were sequentially cultured with increasing concentrations of sertraline for 48 h and apoptotic cells were visualized by caspase-3 and caspase-7 apoptosis assay. **e**, Control and XL413-treated cells were sequentially exposed to 10 μM sertraline, and growth curves were measured by Incucyte live-cell assay. Graphs represent mean ± s.d. from three technical replicates. **f**, Control and XL413-treated cells were sequentially treated with vehicle or 10 μM sertraline for 96 h in colony-formation assays. **g**, Control and XL413-treated Huh7 and Hep3B cells were treated with sertraline (10 μM) for 24 h, and western blot analyses of the indicated proteins of the mTOR signalling pathway were performed. **h**, Hep3B and Huh7 cells were treated with 10 μM XL413 for 10 days before sequential treatment with sertraline (10 μM, 24 h), and RNA sequencing was performed. GSEA indicates that the gene set related

to downregulation of mTOR signalling was negatively enriched in liver cancer cells that were sequentially treated with XL413 and sertraline (Methods). **i, j**, Liver cancer cells with *TP53* mutations (SNU449 and PLC/PRF/5) and lung-cancer cell lines with *TP53* mutations (NCI-H358 and PC9) were treated with 10 μM XL413 or vehicle for 5–7 days, and sequentially exposed to increasing concentrations of AZD8055. Apoptotic cells were visualized by caspase-3 and caspase-7 apoptosis assay 96 h after treatment with AZD8055. **k**, Liver cancer cells with *TP53* mutations (Hep3B and Huh7) were treated with 10 μM XL413 or vehicle for 7–10 days. Control cells and XL413-induced senescent cells were plated and exposed to increasing concentrations of the mTORC1 and mTORC2 inhibitor AZD2014. Apoptotic cells were visualized by caspase-3 and caspase-7 apoptosis assay 96 h after treatment with AZD2014. **l**, Liver cancer cell lines with wild-type *TP53* (SK-Hep1 and Huh6) were treated with 10 μM XL413 or vehicle for 5–7 days before exposure to increasing concentrations of AZD8055. Apoptotic cells were visualized by caspase-3 and caspase-7 apoptosis assay 96 h after treatment with AZD8055. **m**, Control cells and XL413-induced senescent cells were treated with AZD2014 for 48 h. Western blot analysis was performed with the indicated antibodies (left) and the levels of phosphorylated S6RP and phosphorylated 4EBP1 were normalized to the total levels of S6RP and 4EBP1, respectively (right); this shows that treatment with AZD2014 leads to strong inhibition of mTOR signalling in XL413-induced senescent cells. For gel source images, see Supplementary Fig. 1. Data in **a–f** are representative of three independent biological experiments. Data in **h–m** are representative of two independent biological experiments.
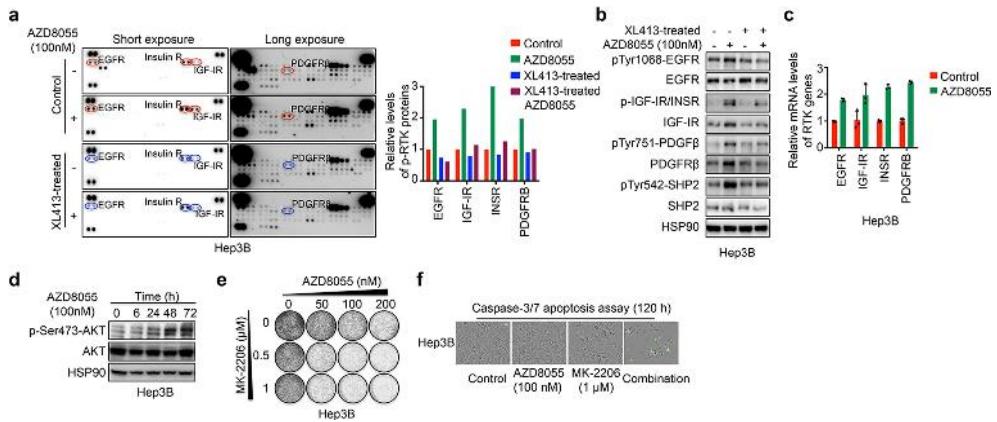
**Extended Data Fig. 7 | The activation of RTK feedback that is induced after treatment with AZD8055 is disrupted in XL413-induced senescent cells. a**, Control cells and XL413-treated Hep3B cells were treated with AZD8055 for 48 h, and extracted proteins were analysed using a human phosphorylated-RTK array kit (left). The levels of phosphorylated RTK proteins were normalized to positive controls (right). **b**, The activation of RTKs identified by RTK arrays and the phosphorylation of SHP2 were validated by western blot analyses. **c**, Hep3B cells were treated with AZD8055 for 48 h before extraction of mRNA, and quantification of the indicated genes for RTK proteins was performed by qRT–PCR. Graph represents mean ± s.d. from three technical replicates. **d**, Hep3B cells were treated with AZD8055, and cell lysates were collected at the indicated time points to perform western blot analyses with the indicated antibodies. **e**, Hep3B cells were exposed to increasing concentrations of the AKT inhibitor MK-2206 in combination with AZD8055, and long-term colony-formation assays were performed; this revealed the synergistic effects of these two compounds on cell viability. **f**, Hep3B cells were treated with AZD8055, MK-2206 or a combination of both compounds at the indicated concentrations for five days, and apoptotic cells were visualized by caspase-3 and caspase-7 apoptosis assay. For gel source images, see Supplementary Fig. 1. All experiments shown (except for the RTK array analyses) are representative of two independent biological experiments.
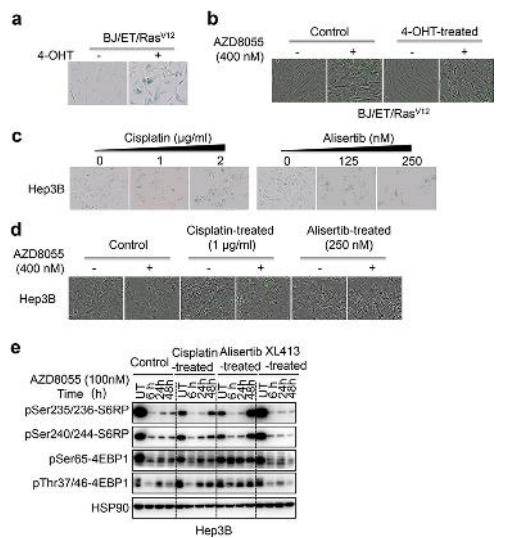
**Extended Data Fig. 8 | Treatment with AZD8055 does not induce apoptosis in cisplatin- or alisertib-induced senescent cells. a**, BJ/ET/RAS[V12] cells were treated with 100 nM 4-OHT for 21 days to induce senescence, as detected by SA-β-gal staining. **b**, Control or senescent BJ/ET/RAS[V12] cells were treated either with vehicle or with 400 nM AZD8055 for 96 h, and apoptotic cells were visualized by caspase-3 and caspase-7 apoptosis assay. **c**, Hep3B cells were cultured in the presence of cisplatin or alisertib (aurora-A kinase inhibitor) for 4 days at the indicated concentrations, and the induction of senescence was detected by SA-β-gal staining. **d**, Hep3B cells were treated with cisplatin (1 μg ml[−1]) or alisertib (250 nM) for 4 days, and subsequently exposed to vehicle or 400 nM AZD8055 for 96 h. Apoptotic cells were visualized by caspase-3 and caspase-7 apoptosis assay. **e**, Control cells, or cisplatin-, alisertib- or XL413-induced senescent cells were treated with AZD8055, and cell lysates were collected at the indicated time points. Western blot analyses were performed with the indicated antibodies, which revealed that the mTOR signalling feedback loop is functional in cisplatin- and alisertib-induced senescent cells (whereas it is efficiently inhibited in XL413-induced senescent cells). For gel source images, see Supplementary Fig. 1. Data in **a**–**e** are representative of two independent biological experiments.
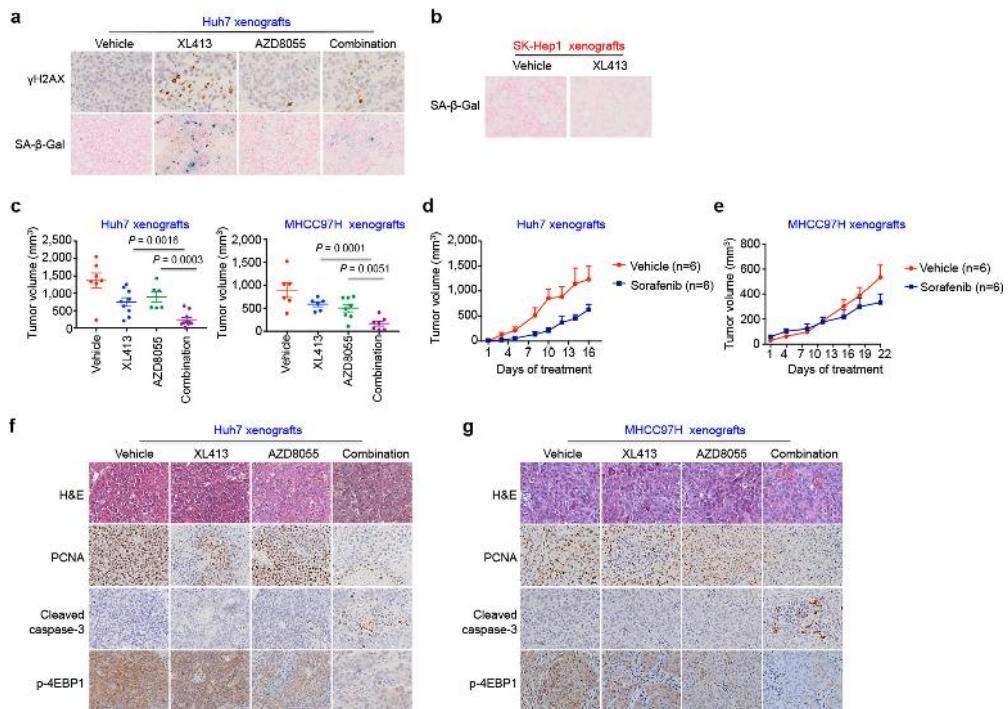
**Extended Data Fig. 9 | Pro-senescence treatment combined with an mTOR inhibitor suppresses tumour growth in liver cancer xenografts. a**, Representative images of γH2AX and SA-β-gal staining performed on formalin-fixed, paraffin-embedded or frozen sections from subcutaneous Huh7-tumour xenografts treated with vehicle, XL413, AZD8055 or combination of both for 12 days. **b**, Representative images of SA-β-gal staining performed on frozen sections from subcutaneous SK-Hep1-tumour xenografts treated either with vehicle or with XL413 for 21 days. **c**, Tumour-volume measurements in mice bearing Huh7- and MHCC97H-tumour xenografts, treated with vehicle, XL413, AZD8055 or a combination of both, at endpoint (12 days for Huh7 and 22 days for MHCC97H). For sample sizes, see Fig. 4a. One mouse in the vehicle group and one mouse in the XL413 group were excluded from the

analysis, because the maximum permitted tumour volumes (2,000 mm³) were reached in these mice before the endpoint of the trial. Graph shows mean ± s.e.m., analysed with two-sided unpaired Student's *t*-test. **d**, **e**, Longitudinal progression of tumour volume in mice bearing Huh7 and MHCC97H tumours, treated with vehicle or sorafenib for 16 or 22 days; this revealed that sorafenib therapy has limited efficacy in these two xenograft models. Graph shows mean ± s.e.m. **f**, **g**, Representative images of haematoxylin and eosin (H & E), PCNA, cleaved caspase-3 and phosphorylated 4EBP1 staining performed on formalin-fixed, paraffin-embedded Huh7 and MHCC97H xenografts from mice killed after the last dose of vehicle, XL413, AZD8055 or a combination of both drugs. Data in **a**, **b**, **f**, **g** are representative of three independent biological experiments.
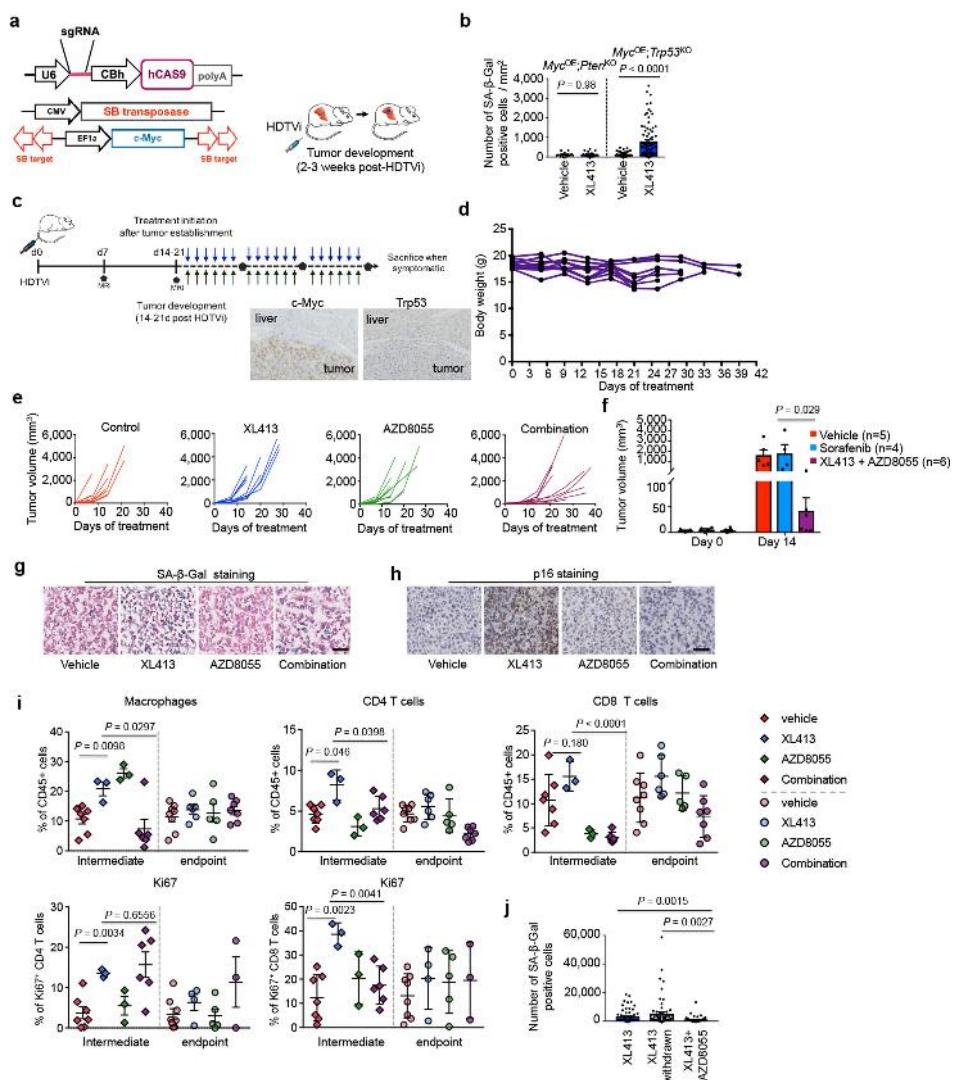
**Extended Data Fig. 10** | See next page for caption.

**Extended Data Fig. 10 | Pro-senescence treatment combined with an mTOR inhibitor suppresses tumour growth in p53-deficient, immunocompetent somatic mouse models of HCC. a**, Schematic of hydrodynamic-tail-vein gene delivery of the *Myc* proto-oncogene transposon system and a CRISPR–Cas9 vector targeting either *Trp53* or the *Pten* tumour suppressor, used to induce models of HCC two to three weeks after HDTVi. **b**, Quantification of SA-β-gal staining performed on frozen sections from mouse models of *Myc*^OE;*Pten*^KO or *Myc*^OE;*Trp53*^KO HCC, 14 days after treatment with vehicle or XL413 monotherapy (results from *Myc*^OE;*Trp53*^KO HCC are also shown in Fig. 4f). For analyses of *Myc*^OE;*Pten*^KO tumours, vehicle-treated, $n = 9$ biologically independent nodules from 3 mice; XL413-treated, $n = 16$ biologically independent nodules from 3 mice. For analyses of *Myc*^OE;*Trp53*^KO tumours, vehicle-treated, $n = 41$ biologically independent nodules from 7 mice; XL413-treated, $n = 81$ biologically independent nodules from 11 mice. Graph shows the mean ± s.e.m. of the number of SA-β-gal+ cells per tumour nodule per mm$^2$. Statistics were calculated by two-sided unpaired Student's *t*-test. **c**, Trial design to evaluate the efficacy of the pro-senescence treatment combined with an mTOR inhibitor in mice bearing *Myc*^OE;*Trp53*^KO HCC. Mice were monitored by weekly MRI after HDTVi, and enrolled into treatments with vehicle, XL413 (100 mg kg$^{-1}$, daily gavage), AZD8055 (20 mg kg$^{-1}$, daily gavage) or XL413 + AZD8055 combination at the first signs of tumour development (revealed by MRI). Drugs were administered six days per week, and mice were killed when they became symptomatic. Immunohistochemical analyses confirmed MYC expression and p53 knockout in endpoint *Myc*^OE;*Trp53*^KO HCC. **d**, Longitudinal individual-body-weight curves from mice bearing *Myc*^OE;*Trp53*^KO tumours, treated with the combination of XL413 + AZD8055. **e**, Individual tumour-growth curves from mice treated with vehicle, XL413, AZD8055 or a combination of both drugs were calculated on the basis of MRI images from mice bearing *Myc*^OE;*Trp53*^KO tumours. **f**, Volumes of *Myc*^OE;*Trp53*^KO tumours from mice

bearing HCC, treated with vehicle ($n = 5$, as shown in Fig. 4c), sorafenib ($n = 4$) or XL413 + AZD8055 ($n = 6$) at day 0 and day 14. Graphs show mean ± s.e.m., analysed with two-sided unpaired Student's *t*-test. **g**, **h**, Representative images of SA-β-gal (**g**) and p16 (**h**) staining performed on frozen and paraffin-embedded sections, respectively, from mice bearing *Myc*^OE;*Trp53*^KO tumours, treated with the indicated drugs and killed at the intermediate time point (14–16 days in time-matched treated cohorts). Quantifications are shown in Fig. 4f, g. Scale bar, 50 μm. **i**, Mice bearing *Myc*^OE;*Trp53*^KO tumours, treated with vehicle, XL413, AZD8055 or a combination of both drugs were killed at the indicated time point after treatment. Tumours were dissociated as single-cell suspensions, and flow cytometry analyses were performed to determine the content of tumour-associated macrophages (CD45+ CD11b+Ly6C−Ly6G−), CD8 T cells (CD45+CD3+CD19−NK1.1−CD8+) and CD4 T cells (CD45+CD3+CD19−NK1.1−CD4+) relative to total CD45+ leucocytes. Cell proliferation (Ki67+) was determined within CD8 T cells and CD4 T cell populations. Graphs show mean ± s.e.m., analysed with two-sided unpaired Student's *t*-test. Sample sizes are given in Methods. **j**, Mice bearing *Myc*^OE;*Trp53*^KO HCC were treated with XL413 ($n = 20$) or XL413 + AZD8055 combination ($n = 8$) for 14 days. Among the XL413-treated mice, a subset ($n = 10$) was killed at 14 days after treatment, concomitantly with the group treated with the combination of drugs. The rest of the XL413-treated mice ($n = 10$) underwent withdrawal of XL413 for 4 days. The absolute number of senescent cells per tumour nodule were visualized by SA-β-gal staining, performed on frozen sections and quantified for each treatment group (XL413-treated, $n = 60$ biologically independent nodules from 10 mice; XL413-withdrawn, $n = 63$ biologically independent nodules from 10 mice; XL413 + AZD8055-treated, $n = 57$ biologically independent nodules from 7 mice). Graphs show mean ± s.e.m. analysed with two-sided unpaired Student's *t*-test. Data in **c** are representative of three independent biological experiments.

# Coordinated alterations in RNA splicing and epigenetic regulation drive leukaemogenesis

Akihide Yoshimi[1,12], Kuan-Ting Lin[2,12], Daniel H. Wiseman[3,4,12], Mohammad Alinoor Rahman[2], Alessandro Pastore[1], Bo Wang[1], Stanley Chun-Wei Lee[1], Jean-Baptiste Micol[5], Xiao Jing Zhang[1], Stephane de Botton[5], Virginie Penard-Lacronique[5], Eytan M. Stein[6], Hana Cho[1], Rachel E. Miles[1], Daichi Inoue[1], Todd R. Albrecht[7], Tim C. P. Somervaille[3], Kiran Batta[4], Fabio Amaral[3], Fabrizio Simeoni[3], Deepti P. Wilks[8], Catherine Cargo[9], Andrew M. Intlekofer[1], Ross L. Levine[1,6], Heidi Dvinge[10], Robert K. Bradley[11], Eric J. Wagner[7], Adrian R. Krainer[2] & Omar Abdel-Wahab[1,6]*

**Transcription and pre-mRNA splicing are key steps in the control of gene expression and mutations in genes regulating each of these processes are common in leukaemia[1,2]. Despite the frequent overlap of mutations affecting epigenetic regulation and splicing in leukaemia, how these processes influence one another to promote leukaemogenesis is not understood and, to our knowledge, there is no functional evidence that mutations in RNA splicing factors initiate leukaemia. Here, through analyses of transcriptomes from 982 patients with acute myeloid leukaemia, we identified frequent overlap of mutations in *IDH2* and *SRSF2* that together promote leukaemogenesis through coordinated effects on the epigenome and RNA splicing. Whereas mutations in either *IDH2* or *SRSF2* imparted distinct splicing changes, co-expression of mutant IDH2 altered the splicing effects of mutant SRSF2 and resulted in more profound splicing changes than either mutation alone. Consistent with this, co-expression of mutant IDH2 and SRSF2 resulted in lethal myelodysplasia with proliferative features in vivo and enhanced self-renewal in a manner not observed with either mutation alone. *IDH2* and *SRSF2* double-mutant cells exhibited aberrant splicing and reduced expression of *INTS3*, a member of the integrator complex[3], concordant with increased stalling of RNA polymerase II (RNAPII). Aberrant *INTS3* splicing contributed to leukaemogenesis in concert with mutant IDH2 and was dependent on mutant SRSF2 binding to *cis* elements in *INTS3* mRNA and increased DNA methylation of *INTS3*. These data identify a pathogenic crosstalk between altered epigenetic state and splicing in a subset of leukaemias, provide functional evidence that mutations in splicing factors drive myeloid malignancy development, and identify spliceosomal changes as a mediator of *IDH2*-mutant leukaemogenesis.**

Mutations in RNA splicing factors are common in cancer and impart specific changes to splicing that are identifiable by mRNA sequencing (RNA-seq)[4–6]. Somatic mutations involving the proline 95 residue of the spliceosome component SRSF2 are among the most recurrent in myeloid malignancies and alter SRSF2 binding to RNA in a sequence-specific manner[6,7]. We analysed RNA-seq data in The Cancer Genome Atlas (TCGA)[1] from 179 patients with acute myeloid leukaemia (AML), evaluating them for spliceosomal alterations. Aberrant splicing events characteristic of *SRSF2* mutations, including *EZH2*[6,7] poison exon inclusion, were observed in 19 patients ($P = 1.6 \times 10^{-12}$; Fisher's exact test; Fig. 1a, Extended Data Fig. 1a, b, Supplementary Table 1). Although only one patient with a mutation in *SRSF2* was reported in the TCGA AML publication[1], mutational analysis of RNA-seq data identified *SRSF2* hotspot mutations in each of these 19 patients

(11% of the patients with AML). Therefore, these data retrospectively identify *SRSF2* as one of the most commonly mutated genes in the TCGA AML cohort.

Notably, 47% of patients with mutated *SRSF2* also had a mutation in *IDH2* and conversely, 56% of patients with mutated *IDH2* had a mutation in *SRSF2* ($P = 1.7 \times 10^{-6}$; Fisher's exact test; Fig. 1b, Extended Data Fig. 1c, d, Supplementary Table 2). Similar results were seen in RNA-seq data from 498 and 263 patients with AML from the Beat AML[8] and Leucegene[9] studies, respectively (Fig. 1c, d, Extended Data Fig. 1e–j, Supplementary Table 2). Across these datasets, variant allele frequencies of *IDH2* and *SRSF2* mutations were high and significantly correlated (Extended Data Fig. 1k), suggesting their common placement as early events in AML.

Beyond these datasets, combined *IDH2* and *SRSF2* mutations were identified in 5.2–6.2% of 1,643 unselected consecutive patients with AML in clinical practice (Supplementary Table 3). Although not statistically significant, *IDH2* and *SRSF2* double-mutant AML cases had the shortest overall survival across the four studied genotypes (Extended Data Fig. 2a). Whereas patients with *IDH2* and *SRSF2* double mutations had mostly intermediate cytogenetic risk, their prognosis was comparable to those with adverse cytogenetic risk (Extended Data Fig. 2b). The patients with *IDH2* and *SRSF2* double mutations were also significantly older than those with mutations in *IDH2* only, or with wild-type *IDH2* and *SRSF2* (Extended Data Fig. 2d; clinical and genetic features are summarized in Extended Data Fig. 2 and Supplementary Table 3).

Mutations in *IDH2* confer neomorphic enzymatic activity that results in the generation of 2-hydroxyglutarate (2HG)[10], which in turn induces DNA hypermethylation via the competitive inhibition of the α-ketoglutarate-dependent enzymes TET1–TET3. Unsupervised hierarchical clustering of DNA methylation data from the TCGA AML cohort revealed that *IDH2* and *SRSF2* double-mutant AML cases form a distinct cluster with higher DNA methylation than *IDH2* single-mutant AML cases (Extended Data Fig. 1l–o). Collectively, these data identify *IDH2* and *SRSF2* double-mutant leukaemia as a recurrent genetically defined AML subset with a distinct epigenomic profile.

We next sought to understand the basis for co-enrichment of *IDH2* and *SRSF2* mutations. Although mutations in splicing factors are frequently found in leukaemias, there is no functional evidence that they can transform cells in vivo. Overexpression of human *IDH2*^*R140Q* or *IDH2*^*R172K* in bone marrow (BM) cells from *Vav-cre Srsf2*^*P95H/+* or *Vav-cre Srsf2*^*+/+* mice revealed a clear collaborative effect between mutant IDH2 and *Srsf2* (Extended Data Fig. 3a). Four weeks after transplantation, the peripheral blood of recipient mice transplanted with *IDH2* and

**Fig. 1 | Frequently co-occurring *IDH2* and *SRSF2* mutations in AML.**
**a**, Heat map of per cent-spliced-in values for mutant *SRSF2*-specific splicing events in TCGA AML samples. **b–d**, Co-occurrence of mutations in *IDH1*, *IDH2*, *TET2* and RNA-splicing factors in the TCGA (**b**), Beat AML (**c**) and Leucegene (**d**) cohorts. Number of patients indicated; co-occurrence or exclusivity noted by colour coding; two-sided Fisher's exact test. Double refers to *SRSF2* and *IDH2* double mutant throughout.

*Srsf2* double-mutant cells had a substantially higher percentage of GFP⁺ cells than in an *Srsf2* wild-type background (Fig. 2a, Extended Data Fig. 3b, c). Moreover, these mice exhibited significant myeloid skewing, macrocytic anaemia and thrombocytopaenia of greater magnitude than seen with mutant *IDH2* (Extended Data Fig. 3d–h). *IDH2* and *Srsf2* double mutants showed no difference in plasma 2HG levels from *IDH2* single mutants (Extended Data Fig. 3i, j). Serial replating of BM cells from leukaemic mice revealed markedly enhanced clonogenicity of *IDH2* and *Srsf2* double-mutant cells compared with other genotypes; the *IDH2* and *Srsf2* cells exhibited a blastic morphology and immature immunophenotype (Extended Data Fig. 3k–m). Consistent with these in vitro results, mice transplanted with *IDH2* and *Srsf2* double-mutant cells developed a lethal myelodysplastic syndrome (MDS) characterized by pancytopaenia, macrocytosis, myeloid dysplasia, expansion of immature BM progenitors and splenomegaly (Fig. 2b, Extended Data Fig. 3n–w). The *IDH2* and *Srsf2* double-mutant cells were also serially transplantable in sublethally irradiated recipients (Fig. 2c, Extended Data Fig. 3x). By contrast, *IDH2* single-mutant controls developed leukocytosis, myeloid skewing without clear dysplasia and less pronounced splenomegaly, whereas *Srsf2* single-mutant cells exhibited impaired repopulation capacity. These results provide evidence that spliceosomal gene mutations can promote leukaemogenesis in vivo.

We next sought to verify the effects of mutant *Idh2* and *Srsf2* using models in which both mutants were expressed from endogenous loci. *Mx1-cre Srsf2^{P95H/+}* mice were crossed with *Idh2^{R140Q/+}* mice to generate control, *Idh2^{R140Q}* single-mutant, *Srsf2^{P95H}* single-mutant and *Idh2* and *Srsf2* double knock-in (DKI) mice (Extended Data Fig. 4a). As expected, 2HG levels in peripheral blood mononuclear cells were increased and 5-hydroxymethylcytosine levels in KIT⁺ BM cells were decreased in *Idh2* single-mutant and DKI primary mice compared with controls (Extended Data Fig. 4b, c). We next performed non-competitive transplantation, in which each mutation was induced alone or together following stable engraftment in recipients. DKI mice showed stable engraftment over time, similar to *Idh2* single-mutant or control mice (Extended Data Fig. 4d). However, DKI mice developed a lethal MDS with proliferative features and significantly shorter survival compared to controls (Fig. 2d). In competitive transplantation, expression of mutant *Idh2^{R140Q}* rescued the impaired self-renewal capacity of *Srsf2* single-mutant cells (Fig. 2e). These observations were supported by an increase in haematopoietic stem–progenitor cells in DKI mice compared with *Srsf2* single-mutant or control mice in primary and serial transplantation (Extended Data Fig. 4e–i). These results confirm cooperativity between mutant IDH2 and SRSF2 in promoting leukaemogenesis in vivo.

On the basis of data identifying 2HG-mediated inhibition of TET2 as a mechanism of *IDH2* mutant leukaemogenesis[11], we also evaluated whether loss of *TET2* might promote transformation of *SRSF2*



**Fig. 2 | Mutant *Idh2* cooperates with mutant *Srsf2* to promote leukaemogenesis. a**, Chimerism of GFP⁺ cells in the blood of recipients transplanted with BM cells with indicated genotypes over time ($n = 5$ per group; data at 0 week represent transduction efficiency; mean percentage ± s.d.; two-way analysis of variance (ANOVA) with Tukey's multiple comparison test). **b–d**, Kaplan–Meier survival analysis of primary recipients (**b**) ($n = 10$ mice per genotype), recipients of serial transplant (**c**) ($n = 5$) and primary recipients transplanted non-competitively with BM cells from knock-in mice (**d**) ($n = 10$). log-rank Mantel–Cox test (two-sided). **e**, Chimerism of peripheral blood CD45.2⁺ cells in competitive transplantation. $n = 10$ mice per group; mean ± s.d.; two-way ANOVA with Tukey's multiple comparison test.

mutant cells. However, deletion of *Tet2* in an *Srsf2* mutant background was insufficient to rescue the impaired self-renewal capacity of *Srsf2* single-mutant cells (Extended Data Fig. 4j–n). Similarly, restoration of TET2 function did not affect the self-renewal capacity of *Idh2* and *Srsf2* double-mutant cells in vivo (Extended Data Fig. 4o–q). These data indicated that the collaborative effects of mutant *Idh2* and *Srsf2* are not solely dependent on TET2. Consistent with this, combined silencing of *Tet2* and *Tet3* partially rescued the impaired replating capacity of *Srsf2* mutant cells in vitro (Extended Data Fig. 4r, s) and the impaired self-renewal of *Srsf2* mutant cells in vivo (Extended Data Fig. 4t–v). Because FTO and ALKBH5—which have roles in RNA processing as $N$6-methyladenosine (m⁶A) RNA demethylases[12,13]—are also dependent on α-ketoglutarate, we investigated the effects of their loss on cooperativity with mutant *Srsf2*. However, collaborative effects were not observed between loss of *Fto* or *Alkbh5* and *Srsf2^{P95H}* (Extended Data Fig. 4w, x).

To understand the basis for cooperation between *IDH2* and *SRSF2* mutations, we next analysed RNA-seq data from the TCGA ($n = 179$ patients), Beat AML ($n = 498$ patients) and Leucegene ($n = 263$ patients) cohorts as well as two previously unpublished RNA-seq datasets targeting defined *IDH2* and *SRSF2* genotype combinations ($n = 42$ patients) and the knock-in mouse models. This revealed that cells with mutations in both *IDH2* and *SRSF2* consistently contained more aberrant splicing events than cells with mutations in *SRSF2* only. Moreover, *IDH2* mutations were associated with a small but reproducible change in RNA splicing (Fig. 3a, b, Extended Data Fig. 5a–g, Supplementary Tables 4–20). By contrast, AML cases in which both *TET2* and *SRSF2* were mutated had fewer changes in splicing than those in which *IDH2* and *SRSF2* were mutated (Extended Data Fig. 5h–m, Supplementary Tables 21, 22).

The majority of splicing changes associated with *SRSF2* mutations involved altered cassette-exon splicing, consistent with *SRSF2* mutations promoting inclusion of C-rich RNA sequences[6,7]. The sequence specificity of mutant SRSF2 on splicing was not influenced by concomitant *IDH2* mutations (Extended Data Fig. 5n–q) and a number of these events were validated by PCR with reverse transcription
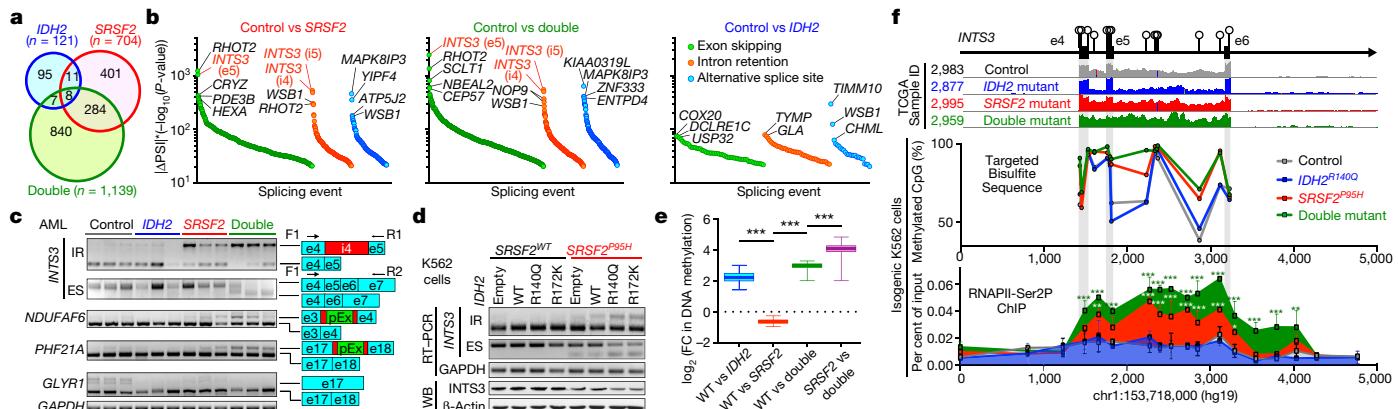
**Fig. 3 | Collaborative effects of mutant *IDH2* and *SRSF2* on aberrant splicing. a**, Venn diagram showing numbers of differentially spliced events from TCGA AML samples. **b**, Number of differentially spliced events ($|\Delta PSI| > 10\%$ and $P < 0.01$) in indicated genotypes are ranked by $((|\Delta PSI| \times |(-\log(P))$ according to class of event (e5, exon 5; i4 and i5, intron 4 and 5, respectively). PSI and $P$ values adjusted for multiple comparisons were calculated using PSI-Sigma[25]. **c**, Representative RT–PCR results of aberrantly spliced transcripts in samples from patients with AML (pEx: exon with premature stop codon; $n = 3$ patients per genotype; three technical replicates with similar results). ES, exon skipping; IR, intron retention; F1, R1 and R2 represent primers used for RT–PCR. **d**, RT–PCR and western blots in isogenic K562 human leukaemia cells (representative images from three biologically independent experiments with similar results). **e**, Mean fold change (expressed

in $\log_2$) in DNA cytosine methylation ($y$ axis) at regions of genomic DNA encoding mRNA that undergo differential splicing ($x$ axis). DNA methylation levels were determined by enhanced reduced representation bisulfite sequencing (eRRBS). $n = 3$ per genotype; the line represents mean, box edges represent the 25th and 75th percentiles and whiskers show 2.5th and 97.5th percentiles; one-way ANOVA with Tukey's multiple comparison test; ***$P < 2.2 \times 10^{-16}$. **f**, Genomic locus of *INTS3* around exons 4–6 with CpG dinucleotides indicated (top), representative RNA-seq from four patients with AML (top; $n = 1$ per genotype), results of targeted bisulfite sequencing (middle; $n = 1$ per genotype) and results of ChIP-walking experiments targeting RNAPII phosphorylated on Ser2 (Ser2P) (bottom; $n = 3$; mean $\pm$ s.d.; two-way ANOVA with Tukey's multiple comparison test compared with control.). Double mutant, $IDH2^{R140Q}SRSF2^{P95H}$. *$P < 0.05$, **$P < 0.01$ and ***$P < 0.001$.

(RT–PCR) of primary AML samples from an independent cohort (Fig. 3c). Among the mis-splicing events in AML with mutations in both *IDH2* and *SRSF2* was a complex event in *INTS3* involving intron retention across two contiguous introns and skipping of the intervening exon (Fig. 3b, c, Extended Data Figs. 5e–g, r–v, 6a–c). Aberrant *INTS3* splicing was demonstrated in isogenic and non-isogenic leukaemia cells with or without *IDH2* and/or *SRSF2* mutations (Fig. 3d, Extended Data Fig. 6d–f), and *INTS3* transcripts with both intron retention and exon skipping resulted in nonsense-mediated decay (Extended Data Fig. 6g–j). Consistent with these observations, INTS3 protein expression was reduced in *SRSF2* mutant cells (Fig. 3d, Extended Data Fig. 6e, f, k–n, Supplementary Table 23). Moreover, silencing of INTS3 was associated with reduced protein levels of additional integrator subunits in *SRSF2* mutant AML compared to *SRSF2* wild-type AML. Consistent with these observations, steady-state protein expression levels of integrator subunits were correlated with one another (Extended Data Fig. 6o). Overall, these data indicate that aberrant splicing and consequent loss of INTS3 was a consistent feature of *IDH2* and *SRSF2* double-mutant cells and was associated with reduced expression of multiple integrator subunits.

We next sought to understand how *IDH2* mutations, which affect the epigenome, might influence splicing catalysis. Splice-site choice is influenced by *cis*-regulatory elements engaged by RNA-binding proteins as well as RNAPII elongation, which is regulated by DNA cytosine methylation and histone modifications[14]. We therefore generated a controlled system to dissect the contribution of RNA-binding elements and DNA methylation to *INTS3* intron retention. We constructed a minigene of *INTS3* spanning exons 4 and 5 and the intervening intron 4 (Extended Data Fig. 7a–c). Transfection of this minigene into leukaemia cells containing combinations of *IDH2* and *SRSF2* mutations revealed that retention of *INTS3* intron 4 is driven by mutant SRSF2 and further enhanced in the *IDH2* and *SRSF2* double-mutant setting (Extended Data Fig. 7d). SRSF2 normally binds C- or G-rich motif sequences in RNA equally well to promote splicing[15]. Leukaemia-associated mutations in *SRSF2* promote its avidity for C-rich sequences while reducing the ability to recognize G-rich sequences[6,7]. Of note, exon 4 of *INTS3* contains the highest number of predicted SRSF2-binding motifs over the entire *INTS3* genomic region (Extended Data Fig. 7c). We evaluated

the role of putative SRSF2 motifs in regulating *INTS3* splicing by mutating all six CCNG motifs in exon 4 to G-rich sequences. In this G-rich version of the minigene, intron retention no longer occurred (*INTS3*-GGNG) (Extended Data Fig. 7e). Conversely, when all G-rich SRSF2 motifs were converted to C-rich sequences (*INTS3*-CCNG), intron retention became evident (Extended Data Fig. 7f). These results confirmed the sequence-specific activity of mutant SRSF2 in *INTS3* intron retention and identified a role for mutant IDH2 in regulating splicing.

Because *IDH2* mutations promote increased DNA methylation and DNA methylation can affect splicing[14], we generated genome-wide maps of DNA cytosine methylation from patients with AML across four genotypes (Supplementary Table 23). This revealed that differentially spliced events in *IDH2* single-mutant as well as *IDH2* and *SRSF2* double-mutant AML (compared to *IDH2* and *SRSF2* wild-type and *SRSF2* single-mutant AML) contained significant hypermethylation of DNA. Thus regions of differential DNA hypermethylation significantly overlapped with regions of differential RNA splicing (Fig. 3e, Extended Data Fig. 7j).

The above results suggest a strong link between increased DNA methylation mediated by mutant IDH2 and altered RNA splicing by mutant SRSF2. To evaluate this further, we next examined DNA methylation levels around endogenous *INTS3* exons 4–6 by targeted bisulfite sequencing. This revealed increased DNA methylation at all CpG dinucleotides in this region in *IDH2* and *SRSF2* double-mutant cells compared to control or single-mutant cells (Fig. 3f, Extended Data Fig. 7k). A functional role of DNA methylation at these sites was verified by evaluating splicing in versions of the *INTS3* minigene in which each CG dinucleotide was converted to an AT to prevent cytosine methylation. In these CG-to-AT versions of the minigene, *IDH2* mutations no longer promoted mutant-*SRSF2*-mediated intron retention (Extended Data Fig. 7g–i). As further confirmation of the influence of mutant IDH2 on *INTS3* splicing, cell-permeable 2HG increased *INTS3* intron retention whereas treatment of *IDH2* and *SRSF2* double-mutant cells with the DNA methyltransferase inhibitor 5-aza-2′-deoxycytidine (5-AZA-CdR) inhibited *INTS3* intron retention (Extended Data Fig. 7l, m).

Given that changes in epigenetic state may affect splicing by influencing RNAPII stalling[14,16], we evaluated the abundance of RNAPII
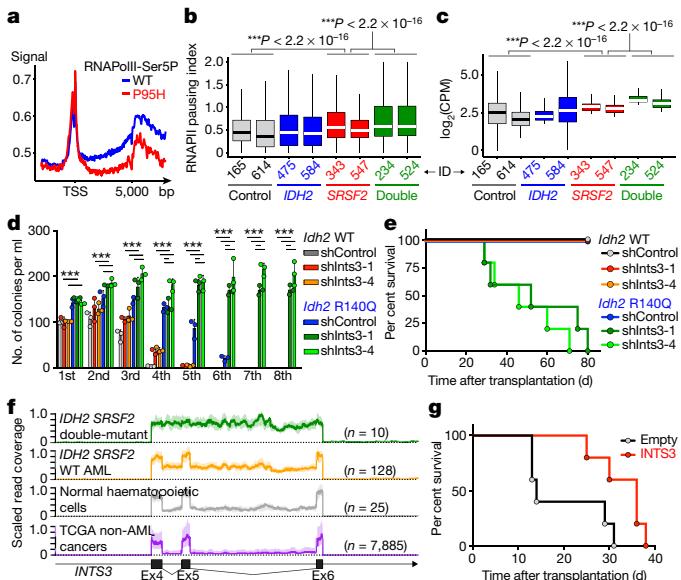
**Fig. 4 | RNAPII stalling in *IDH2* and *SRSF2* double-mutant AML and contribution of INTS3 loss to leukaemogenesis. a**, Metagene plot of genome-wide RNAPII Ser5 phosphorylation (Ser5P) in isogenic *SRSF2*[WT] or *SRSF2*[P95H] mutant K562 cells. **b, c**, RNAPII pausing index[20] in primary AML samples calculated as the ratio of normalized ChIP–seq reads of RNAPII Ser5P on transcription start sites (TSSs) ($\pm$ 250 bp) over that of the corresponding bodies (+500 bp to +1000 bp from TSSs) (**b**) and RNAPII abundance over the differentially spliced regions between *SRSF2* single-mutant and *IDH2* and *SRSF2* double-mutant AML determined by RNAPII Ser2P ChIP–seq (*y* axis shows log$_2$(counts per million (CPM))) (**c**). *x* axis shows patient ID. Box plots were generated from ChIP–seq data from an individual primary AML sample; the line represents mean, box edges represent 25th and 75th percentiles and whiskers show 2.5th and 97.5th percentiles. One-way ANOVA with Tukey's multiple comparison test. **d**, Colony numbers from serial replating assays of either *Mx1-cre Idh2*[+/+] or *Idh2*[R140Q/+] BM cells transduced with shRNA targeting *Ints3* (shInts3-1 or shInts3-4) or control shRNA (shControl). *n* = 3 biologically independent experiments; mean $\pm$ s.d.; two-way ANOVA with Tukey's multiple comparison test. ***$P$ < 0.004. **e, g**, Kaplan–Meier survival analysis of recipients. *n* = 5 per group; log-rank (Mantel–Cox) test (two-sided). In **e**: shInts3-1 versus shControl, $P$ = 0.0018; shInts3-4 versus shControl, $P$ = 0.0018. In **g**, $P$ = 0.034 (INTS3 versus empty vector control). **f**, RNA-seq read coverage between exons 4–6 of *INTS3* and 1,000 bp either side of *INTS3* is scaled and shown as mean (line) $\pm$ s.d. (shaded region) (generated from TCGA datasets; sample list shown in legend for Extended Data Fig. 10e).

using chromatin precipitation with DNA sequencing (ChIP–seq) in isogenic *SRSF2*[WT] and *SRSF2*[P95H] cells as well as the primary samples from patients with AML. This revealed increased promoter-proximal transcriptional pausing and decreased RNAPII occupancy over gene bodies in *SRSF2* mutant cells, which was further enhanced in *IDH2* and *SRSF2* double-mutant cells (Fig. 4a, b, Extended Data Fig. 7n–q, Supplementary Table 23). Transcriptional pausing was also evident at *INTS5* and *INTS14* in *SRSF2* mutant cells (Extended Data Fig. 7r, s), which—in combination with aberrant splicing of several integrator subunits (Supplementary Table 24)—suggested impaired function of the entire integrator complex in *SRSF2* mutant cells. Similar to DNA cytosine methylation levels, RNAPII was more abundant over differentially spliced regions in *SRSF2* single-mutant AML than in *SRSF2* wild-type AML, and further enhanced over differentially spliced regions in *IDH2* and *SRSF2* double-mutant AML compared with those in *SRSF2* single-mutant AML (Fig. 4c, Extended Data Fig. 7t).

The above data provide further links between increased DNA cytosine methylation and RNAPII stalling with altered RNA splicing in *IDH2* and *SRSF2* double-mutant AML. To further evaluate this model, we performed ChIP for RNAPII across 4,766 bp of the *INTS3* locus in isogenic leukaemia cells (Fig. 3f). This revealed

substantial accumulation of RNAPII across this locus in *IDH2* and *SRSF2* double-mutant cells. Treatment with 5-AZA-CdR significantly reduced RNAPII stalling, which was coupled with decreased aberrant *INTS3* splicing (Extended Data Fig. 7k–m). These data reveal that *IDH2* and *SRSF2* mutations coordinately dysregulate splicing through alterations in RNAPII stalling in addition to aberrant sequence recognition of *cis* elements in RNA.

*INTS3* encodes a component of the integrator complex that participates in small nuclear RNA (snRNA) processing[3] in addition to RNAPII pause–release[17]. Consistent with this, *SRSF2* single-mutant cells had altered snRNA cleavage similar to those seen with direct INTS3 downregulation, which was exacerbated in *IDH2* and *SRSF2* double-mutant cells (Extended Data Fig. 8a–h). Attenuation of INTS3 expression in *SRSF2* mutant cells caused a blockade of myeloid differentiation, an effect further enhanced in an *IDH2* mutant background (Extended Data Fig. 8i–n). Notably, direct *Ints3* downregulation in the *Idh2*[R140Q/+] background resulted in enhanced clonogenic capacity of cells with an immature morphology and immunophenotype (Fig. 4d, Extended Data Fig. 8o–r) and promoted clonal dominance of *Idh2* mutant cells (Fig. 4e, Extended Data Fig. 9a–d). Moreover, mice transplanted with *Idh2*[R140Q/+] BM cells treated with short hairpin RNA (shRNA) targeting *Ints3* exhibited myeloid skewing, anaemia and thrombocytopaenia (Extended Data Fig. 9e–g) and developed a lethal MDS with proliferative features—phenotypes resembling those seen in *IDH2* and *Srsf2* double-mutant mice (Fig. 4e, Extended Data Fig. 9h, i).

The defects in snRNA processing in *SRSF2* single-mutant and *IDH2* and *SRSF2* double-mutant cells were partially rescued by INTS3 cDNA expression (Extended Data Fig. 8s–x). In addition, restoration of INTS3 expression released *SRSF2* single-mutant and *IDH2* and *SRSF2* double-mutant HL-60 cells from differentiation block (Extended Data Fig. 8y, z). Xenografts of *IDH2* and *SRSF2* double-mutant HL-60 cells demonstrated that forced expression of INTS3 induced myeloid differentiation and slowed leukaemia progression in vivo (Extended Data Fig. 9j-s). Collectively, these data suggest that INTS3 loss due to aberrant splicing by mutant IDH2 and SRSF2 contributes to leukaemogenesis.

Although loss of INTS3 resulted in measurable changes in snRNA processing, the degree of snRNA mis-processing did not have a substantial effect on splicing as determined by RNA-seq of *IDH2*[R140Q] mutant HL-60 cells with INTS3 silencing. By contrast, INTS3 depletion in these cells significantly affected transcriptional programs associated with myeloid differentiation, multiple oncogenic signalling pathways, RNAPII elongation-linked transcription and DNA repair (Extended Data Fig. 10a–d, Supplementary Table 25). This latter association of INTS3 loss with DNA repair is potentially consistent with previous reports that sensor of single-stranded DNA complexes containing INTS3 participate in DNA damage response[18,19].

These data uncover an important role for RNA splicing alterations in *IDH2* mutant tumorigenesis and identify perturbations in integrator as a driver of transformation of *IDH2* and *SRSF2* mutant cells. However, INTS3 is not known to be recurrently affected by coding-region alterations in leukaemias. We therefore evaluated *INTS3* splicing across 32 additional cancer types as well as normal blood cells to evaluate whether aberrant *INTS3* splicing might be a common mechanism in AML. This revealed that, whereas *INTS3* mis-splicing is most evident in *IDH2* and *SRSF2* double-mutant AML, *INTS3* mis-splicing is also prevalent across other molecular subtypes of AML but is not present in blood cells from healthy subjects or RNA-seq data from more than 7,000 samples from other cancer types (Fig. 4f, Extended Data Fig. 10e, f). To further evaluate the effects of enforced INTS3 expression in myeloid leukaemia with a wild-type splicing phenotype, we used *MLL-AF9;Nras*[G12D] mouse leukaemia (RN2) cells. INTS3 overexpression reduced colony-forming capacity (Extended Data Fig. 10g, h) and enhanced differentiation of RN2 cells, resulting in decelerated leukaemia progression in vivo (Fig. 4g, Extended Data Fig. 10i–s).

These data highlight a role for loss of INTS3 in broad genetic subtypes of AML. Further efforts to determine how integrator loss

promotes leukaemogenesis and other non-mutational mechanisms mediating *INTS3* aberrant splicing will be critical for understanding and targeting leukaemias with integrator loss. Previous studies have identified that integrator[17,20] and SRSF2[21] have direct roles in modulating transcriptional pause–release. The accumulation of RNAPII at certain mis-spliced loci in this study is consistent with recent data that suggest that mutant SRSF2 is defective in promoting RNAPII pause–release[22]. Identifying how aberrant splicing mediated by mutant SRSF2 is influenced by altered RNAPII pause–release may therefore be informative.

In addition to modifying splicing in *SRSF2* mutant cells, *IDH2* mutations were associated with reproducible changes in splicing in haematopoietic cells. There is a strong correlation between aberrant splicing in *IDH2* and *IDH1* mutant low-grade gliomas ($P = 2.2 \times 10^{-16}$ (binominal proportion test); Extended Data Fig. 10t–w, Supplementary Tables 26–28). A significant number of splicing events that were dysregulated in *IDH2* mutant AML from the TCGA and Leucegene cohorts were differentially spliced in *IDH2* mutants versus *IDH1* and *IDH2* wild-type low-grade gliomas ($P = 1.8 \times 10^{-9}$ and $P = 1.3 \times 10^{-8}$, respectively; binominal proportion test). These data suggest that *IDH1* and *IDH2* mutations impart a consistent effect on splicing regardless of tumour type. Finally, these results have important translational implications given the substantial efforts to pharmacologically inhibit mutant IDH1 and IDH2 as well as mutant splicing factors[23,24]. The frequent coexistence of *IDH2* and *SRSF2* mutations underscores the enormous therapeutic potential for modulation of splicing in the approximately 50% of patients with *IDH2* mutant leukaemia who also have a spliceosomal gene mutation.

### Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-019-1618-0.

1. The Cancer Genome Atlas Research Network. Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia. *N. Engl. J. Med.* **368**, 2059–2074 (2013).
2. Papaemmanuil, E. et al. Clinical and biological implications of driver mutations in myelodysplastic syndromes. *Blood* **122**, 3616–3627, quiz 3699 (2013).
3. Wu, Y., Albrecht, T. R., Baillat, D., Wagner, E. J. & Tong, L. Molecular basis for the interaction between Integrator subunits IntS9 and IntS11 and its functional importance. *Proc. Natl Acad. Sci. USA* **114**, 4394–4399 (2017).
4. Darman, R. B. et al. Cancer-associated SF3B1 hotspot mutations induce cryptic 3′ splice site selection through use of a different branch point. *Cell Rep.* **13**, 1033–1045 (2015).
5. Ilagan, J. O. et al. *U2AF1* mutations alter splice site recognition in hematological malignancies. *Genome Res.* **25**, 14–26 (2015).
6. Kim, E. et al. *SRSF2* mutations contribute to myelodysplasia by mutant-specific effects on exon recognition. *Cancer Cell* **27**, 617–630 (2015).
7. Zhang, J. et al. Disease-associated mutation in *SRSF2* misregulates splicing by altering RNA-binding affinities. *Proc. Natl Acad. Sci. USA* **112**, E4726–E4734 (2015).
8. Tyner, J. W. et al. Functional genomic landscape of acute myeloid leukaemia. *Nature* **562**, 526–531 (2018).
9. Lavallée, V. P. et al. The transcriptomic landscape and directed chemical interrogation of *MLL*-rearranged acute myeloid leukemias. *Nat. Genet.* **47**, 1030–1037 (2015).
10. Dang, L. et al. Cancer-associated IDH1 mutations produce 2-hydroxyglutarate. *Nature* **462**, 739–744 (2009).
11. Figueroa, M. E. et al. Leukemic IDH1 and IDH2 mutations result in a hypermethylation phenotype, disrupt TET2 function, and impair hematopoietic differentiation. *Cancer Cell* **18**, 553–567 (2010).
12. Jia, G. et al. N6-methyladenosine in nuclear RNA is a major substrate of the obesity-associated FTO. *Nat. Chem. Biol.* **7**, 885–887 (2011).
13. Zheng, G. et al. ALKBH5 is a mammalian RNA demethylase that impacts RNA metabolism and mouse fertility. *Mol. Cell* **49**, 18–29 (2013).
14. Naftelberg, S., Schor, I. E., Ast, G. & Kornblihtt, A. R. Regulation of alternative splicing through coupling with transcription and chromatin structure. *Annu. Rev. Biochem.* **84**, 165–198 (2015).
15. Daubner, G. M., Cléry, A., Jayne, S., Stevenin, J. & Allain, F. H. A syn-anti conformational difference allows SRSF2 to recognize guanines and cytosines equally well. *EMBO J.* **31**, 162–174 (2012).
16. Shukla, S. et al. CTCF-promoted RNA polymerase II pausing links DNA methylation to splicing. *Nature* **479**, 74–79 (2011).
17. Gardini, A. et al. Integrator regulates transcriptional initiation and pause release following activation. *Mol. Cell* **56**, 128–139 (2014).
18. Huang, J., Gong, Z., Ghosal, G. & Chen, J. SOSS complexes participate in the maintenance of genomic stability. *Mol. Cell* **35**, 384–393 (2009).
19. Li, Y. et al. HSSB1 and hSSB2 form similar multiprotein complexes that participate in DNA damage response. *J. Biol. Chem.* **284**, 23525–23531 (2009).
20. Stadelmayer, B. et al. Integrator complex regulates NELF-mediated RNA polymerase II pause/release and processivity at coding genes. *Nat. Commun.* **5**, 5531 (2014).
21. Ji, X. et al. SR proteins collaborate with 7SK and promoter-associated nascent RNA to release paused polymerase. *Cell* **153**, 855–868 (2013).
22. Chen, L. et al. The augmented R-loop is a unifying mechanism for myelodysplastic syndromes induced by high-risk splicing factor mutations. *Mol. Cell* **69**, 412–425 (2018).
23. Seiler, M. et al. H3B-8800, an orally available small-molecule splicing modulator, induces lethality in spliceosome-mutant cancers. *Nat. Med.* **24**, 497–504 (2018).
24. Stein, E. M. et al. Enasidenib in mutant *IDH2* relapsed or refractory acute myeloid leukemia. *Blood* **130**, 722–731 (2017).

## METHODS

**Data reporting.** The number of mice in each experiment was chosen to provide 90% statistical power with a 5% error level. Otherwise, no statistical methods were used to predetermine sample size. The experiments were not randomized. The investigators were not blinded to allocation during experiments and outcome assessment.

**Mice.** All mice were housed at Memorial Sloan Kettering Cancer Center (MSK). All mouse procedures were completed in accordance with the Guidelines for the Care and Use of Laboratory Animals and were approved by the Institutional Animal Care and Use Committees at MSK. Six- to eight-week-old week female CD45.1 C57BL/6 mice were purchased from The Jackson Laboratory (Stock No: 002014) . Male and female CD45.2 $Srsf2^{P95H/+}$ conditional knock-in mice, $Idh2^{R140Q/+}$ conditional knock-in mice, and $Tet2$ conditional knockout mice (all on C57BL/6 background) were also analysed and used as bone marrow donors (generation of these mice were as described[6,26,27]). For BM transplantation assays with IDH2 overexpression, $Srsf2^{P95H/+}$ and littermate control mice were crossed to $Vav-cre$ transgenic mice[28]. CBC analysis was performed on peripheral blood collected from submandibular bleeding, using a Procyte Dx Hematology Analyzer (IDEXX Veterinary Diagnostics). For all mouse experiments, the mice were monitored closely for signs of disease or morbidity daily and were euthanized for visible tumour formation at tumour volume >1 cm³, failure to thrive, weight loss >10% total body weight, open skin lesions, bleeding, or any signs of infection. In none of the experiments were these limits exceeded.

**BM transplantation assays.** Freshly dissected femurs and tibias were isolated from $Mx1-cre$, $Mx1-cre/Idh2^{R140Q/+}$, $Mx1-cre\ Srsf2^{P95H/+}$, $Mx1-cre\ Idh2^{R140Q/+}Srsf2^{P95H/+}$, $Mx1-cre\ Tet2^{fl/fl}$, or $Mx1-cre\ Tet2^{fl/fl}Srsf2^{P95H/+}$ CD45.2$^+$ mice. BM was flushed with a 3-cm³ insulin syringe into cold PBS supplemented with 2% bovine serum albumin to generate single-cell suspensions. BM cells were pelleted by centrifugation at 1,500 r.p.m. for 4 min and red blood cells (RBCs) were lysed in ammonium chloride-potassium bicarbonate lysis (ACK) buffer (150 mM NH$_4$Cl + 10 mM KHCO$_3$ + 0.1 mM EDTA; Thermo Fisher Scientific) for 3 min on ice. After centrifugation, cells were resuspended in PBS/2% BSA, passed through a 40-μm cell strainer, and counted. For competitive transplantation experiments, $0.5 \times 10^6$ BM cells from $Mx1-cre$, $Mx1-cre\ Idh2^{R140Q/+}$, $Mx1-cre\ Srsf2^{P95H/+}$, $Mx1-cre\ Idh2^{R140Q/+}Srsf2^{P95H/+}$, $Mx1-cre\ Tet2^{fl/fl}$ or $Mx1-cre\ Tet2^{fl/fl}Srsf2^{P95H/+}$ CD45.2$^+$ mice were mixed with $0.5 \times 10^6$ wild-type (WT) CD45.1$^+$ BM and transplanted via tail-vein injection into 8-week-old lethally irradiated (900 cGy) CD45.1$^+$ recipient mice. The CD45.1$^+$:CD45.2$^+$ ratio was confirmed to be approximately 1:1 by flow cytometry analysis pre-transplant. To activate the conditional alleles, mice were treated with 3 doses of polyinosinic:polycytidylic acid (pIpC; 12 mg per kg (body weight) per day; GE Healthcare) every second day via intraperitoneal injection. Peripheral blood chimerism was assessed every four weeks by flow cytometry. For noncompetitive transplantation experiments, $1 \times 10^6$ total BM cells from $Mx1-cre$, $Mx1-cre\ Idh2^{R140Q/+}$, $Mx1-cre\ Srsf2^{P95H/+}$, $Mx1-cre\ Idh2^{R140Q/+}Srsf2^{P95H/+}$, $Mx1-cre\ Tet2^{fl/fl}$, or $Mx1-cre\ Tet2^{fl/fl}Srsf2^{P95H/+}$ CD45.2$^+$ mice were injected into lethally irradiated (950 cGy) CD45.1$^+$ recipient mice. Peripheral blood chimerism was assessed as described for competitive transplantation experiments. Additionally, for each bleeding whole blood cell counts were measured on an automated blood analyser. Mice that were lost owing to pIpC toxicity were excluded from analysis.

**Retroviral transduction and transplantation of primary haematopoietic cells.** $Vav-cre\ Srsf2^{+/+}$ and $Vav-cre\ Srsf2^{P95H/+}$ mice were treated with a single dose of 5-fluorouracil (150 mg kg$^{-1}$) followed by BM collection from the femurs, tibias and pelvic bones 5 days later. RBCs were removed by ACK lysis buffer, and nucleated BM cells were transduced with viral supernatants containing MSCV-$IDH2^{WT/R140Q/R172K}$-IRES-GFP for 2 days in RPMI/20% FCS supplemented with mouse stem cell factor (mSCF, 25 ng ml$^{-1}$), mouse interleukin-3 (mIL3, 10 ng ml$^{-1}$) and mIL6 (10 ng ml$^{-1}$), followed by injection of about $0.5 \times 10^6$ cells per recipient mouse via tail vein injection into lethally irradiated (950 cGy) CD45.1$^+$ mice. Transplantation of primary BM cells with TET2 catalytic domain cDNA and anti-$Ints3$ or $Tet3$ shRNAs was similarly performed. For secondary transplantation experiments, 8-week old, lethally (900–950 cGy) or sub-lethally (450–700 cGy) irradiated C57/BL6 recipient mice were injected with unsorted $1 \times 10^6$ BM cells from the primary transplantation. $IDH2^{WT} + Srsf2^{WT}$ and $IDH2^{WT} + Srsf2^{P95H}$ mice were euthanized at day 315 post-transplant to collect BM for the serial transplantation. All cytokines were purchased from R&D Systems.

**Flow cytometry analyses and antibodies.** Surface-marker staining of haematopoietic cells was performed by first lysing cells with ACK lysis buffer and washing cells with ice-cold PBS. Cells were stained with antibodies in PBS/2% BSA for 30 min on ice. For haematopoietic stem/progenitor staining, cells were stained with the following antibodies: B220-APCCy7 (clone: RA3-6B2; purchased from BioLegend; catalogue no.: 103224; dilution: 1:200); B220-Bv711 (RA3-6B2; BioLegend; 103255; 1:200); CD3-PerCPCy5.5 (17A2; BioLegend; 100208; 1:200); CD3-APC (17A2; BioLegend; 100236; 1:200); CD3-APCCy7 (17A2; BioLegend; 100222; 1:200); Gr1-PECy7 (RB6-8C5; eBioscience; 25-5931-82; 1:500); CD11b-PE

(M1/70; eBioscience; 12-0112-85; 1:500); CD11b-APCCy7 (M1/70; BioLegend; 101226; 1:200); CD11c-APCCy7 (N418; BioLegend; 117323; 1:200); NK1.1-APCCy7 (PK136; BioLegend; 108724; 1:200); Ter119-APCCy7 (BioLegend; 116223: 1:200); KIT-APC (2B8; BioLegend; 105812; 1:200); KIT-PerCPCy5.5 (2B8; BioLegend; 105824; 1:100); KIT-Bv605 (ACK2; BioLegend; 135120; 1:200); Sca1-PECy7 (D7; BioLegend; 108102; 1:200); CD16/CD32 (FcγRII/III)-Alexa700 (93; eBioscience; 56-0161-82; 1:200); CD34-FITC (RAM34; BD Biosciences; 553731; 1:200); CD45.1-FITC (A20; BioLegend; 110706; 1:200); CD45.1-PerCPCy5.5 (A20; BioLegend; 110728; 1:200); CD45.1-PE (A20; BioLegend; 110708; 1:200); CD45.1-APC (A20; BioLegend; 110714; 1:200); CD45.2-PE (104; eBioscience; 12-0454-82; 1:200); CD45.2-Alexa700 (104; BioLegend; 109822; 1:200); CD45.2-Bv605 (104; BioLegend; 109841; 1:200); CD48-Bv711 (HM48-1; BioLegend; 103439; 1:200); CD150 (9D1; eBioscience; 12-1501-82; 1:200). DAPI was used to exclude dead cells. For sorting human leukaemia cells, cells were stained with a lineage cocktail including CD34-PerCP (8G12; BD Biosciences; 345803; 1:200); CD117-PECy7 (104D2; eBioscience; 25-1178-42; 1:200); CD33-APC (P67.6; BioLegend; 366606; 1:200); HLA-DR-FITC (L243; BioLegend; 307604; 1:200); CD13-PE (L138; BD Biosciences; 347406; 1:200); CD45-APC-H7 (2D1; BD Biosciences; 560178; 1:200). The composition of mature haematopoietic cell lineages in the BM, spleen and peripheral blood was assessed using a combination of CD11b, Gr1, B220, and CD3. For the haematopoietic stem and progenitor analysis, a combination of CD11b, CD11c, Gr1, B220, CD3, NK1.1 and Ter119 was stained as lineage-positive cells. Fluorescence-activated cell sorting (FACS) was performed on a FACS Aria, and analysis was performed on an LSRII or LSR Fortessa (BD Biosciences). For western blotting, DNA dot blot assays, and ChIP assays, the following antibodies were used: INTS1 (purchased from Bethyl laboratories; catalogue no.: A300-361A; dilution: 1:1,000), INTS2 (Abcam; ab74982; 1:1,000), INTS3 (Bethyl laboratories; A300-427A; 1:1,000, Abcam; ab70451; 1:1,000), INTS4 (Bethyl laboratories; A301-296A; 1:1,000), INTS5 (Abcam; ab74405; 1:1,000), INTS6 (Abcam; ab57069; 1:1,000), INTS7 (Bethyl laboratories; A300-271A; 1:1,000), INTS8 (Bethyl laboratories; A300-269A; 1:1,000), INTS9 (Bethyl laboratories; A300-412A; 1:1,000), INTS11 (Abcam; ab84719; 1:1,000), Flag-M2 (Sigma-Aldrich; F-1084; 1:1,000), Myc-tag (Cell Signaling; 2276S; 1:1,000), β-actin (Sigma-Aldrich; A-5441; 1:2,000), 5-hydroxymehylcytosine (5hmC) (Active motif; 39769), RNAPII CTD repeat YSPTSPS (phospho S2) (Abcam; ab5095), RNAPII CTD repeat YSPTSPS (phospho S5) (Abcam; ab5408), and UPF1 (Abcam; ab109363; 1:1,000).

**Minigene assay.** We constructed $INTS3$-WT minigene spanning exons 4 to 5 of human $INTS3$ into pcDNA3.1(+) vector (Invitrogen) using BamHI and XhoI sites, respectively. Artificial mutations were engineered into $INTS3$-WT minigene using the QuikChange Site-Directed Mutagenesis Kit (Agilent) to generate $INTS3$-GGNG, $INTS3$-CCNG, $INTS3$-WT_CG(−) $INTS3$-GGNG_CG(−), and $INTS3$-CCNG_CG(−) minigenes, respectively, and the sequences of inserts were verified by Sanger sequencing. Plasmids (1 μg) were transfected using Lipofectamine LTX reagent with PLUS reagent (Invitrogen) including 0.2 μg of eGFP and 0.8 μg of $INTS3$ minigene, per well of a 6-well plate. Total RNA was extracted 48 h after transfection using TRIzol reagent (Ambion), followed by DNase I treatment (Qiagen). cDNA was synthesized with an oligo-dT primer using ImProm-II reverse transcriptase (Promega). Radioactive PCR was done with $^{32}$P-α-dCTP, 1.25 units of AmpliTaq (Invitrogen) and 26 cycles using primer pairs 5′-GCTTGGTACCGAGCTCGGATC-3′ (vector-specific forward primer) and 5′-CAGTTCCCGTACCAACCACAC-3′ (reverse primer for $INTS3$ versions of minigene), or 5′-CAGTTCCATTACCAACCACAC-3′ (reverse primer for $INTS3$_CG(-) versions of minigene). Products were run on a 5% PAGE and the bands were quantified using a Typhoon FLA 7000 (GE Healthcare). eGFP was used as a control for transfection efficiency and exogenous eGFP was amplified using a vector specific forward primer and reverse primer on eGFP. eGFP products were loaded after we ran the INTS3 products for 20–30 min. Percentages of intron 4 retention were normalized against exogenous eGFP.

**Cell culture.** K562 (human chronic myeloid/erythroleukaemia cell line) and HL-60 (human promyelocytic leukaemia cell line) leukaemia cells, K052 (human multilineage leukaemia cell line) leukaemia cells, TF1 (human erythroleukaemia cell line) leukaemia cells, $MLL-AF9/Nras^{G12D}$ murine leukaemia (RN2) cells[29], and Ba/F3 (murine pro-B cell line) cells were cultured in RPMI/10% fetal calf serum (FCS, heat inactivated), RPMI/20% FCS, RPMI/10% FCS + human granulocyte-macrophage colony-stimulating factor (GM-CSF, R&D Systems; 5 ng ml$^{-1}$), and RPMI/10% FCS + mIL3 (R&D Systems; 1 ng ml$^{-1}$), respectively. None of the cell lines above were listed in the database of commonly misidentified cell lines maintained by ICLAC and NCBI Biosample.

MSCV-$IDH2^{WT/R140Q/R172K}$-IRES-GFP, MSCV-3×Flag-$INTS3$-puro, MSCV-IRES-3×Flag-$INTS3$-mCherry, MSCV-IRES-TET2 catalytic domain cDNA-mCherry ('TET2CD'), and empty vectors of these constructs were used for retroviral overexpression studies and pRRLSIN.cPPT.PGK-mCherry.WPRE-$SRSF2^{WT/P95H}$ constructs were used for lentiviral overexpression studies. TET2CD

cDNA fragment with Myc tag was generated by PCR amplification using pCMVT$_N$T- TET2CD[30] as a template and inserted in the BglII restriction sites of MSCV-IRES-mCherry. Retroviral supernatants were produced by transfecting 293 GPII cells with cDNA constructs and the packaging plasmid VSV.G using XtremeGene9 (Roche) or polyethylenimine hydrochloride (Polysciences). Lentiviral supernatants were produced by similarly transfecting HEK293T cells with cDNA constructs and the packaging plasmid VSV.G and psPAX2. Virus supernatants were used for transduction in the presence of polybrene (5 μg ml$^{-1}$). GFP$^+$mCherry$^+$ double-positive HL-60 cells and mCherry$^+$ K562 cells were FACS-sorted to obtain cells expressing wild-type or mutant IDH2 and SRSF2 in various combinations. Isogenic HL-60 cells transduced with 3×Flag-tagged INTS3 or empty vector were obtained by puromycin selection (1 μg ml$^{-1}$). To let the cells fully establish epigenetic changes, they were analysed after culture for more than 30 days.

For in vitro colony-forming assays, a single-cell suspension was prepared and 15,000 cells per 1.5 ml were plated in triplicates in cytokine-supplemented methylcellulose medium (MethoCult GF M3434; StemCell Technologies), and colonies were enumerated every week. For the colony-forming assays shown in Extended Data Fig. 3k, $IDH2^{WT} + Srsf2^{WT}$ and $IDH2^{WT} + Srsf2^{P95H}$ mice were euthanized at day 315 post-transplant to collect BM as controls.

**shRNA-mediated silencing.** shRNAs against human *INTS3* (hINTS3), mouse *Ints3* (mInts3), and mouse *Tet3* (mTet3) were cloned into MLS-E-Cherry and/or MLS-E-GFP vector and those against human *UPF1* (hUPF1), mouse *Fto* (mFto), and mouse *Alkbh5* (mAlkbh5) were cloned into LT3GEPIR (pRRL) Lenti-GFP-Puro-Tet-ON all-in-one vector. The antisense sequences were: hINTS3-1: TTTTCGAAACATAACCAGGTTA; hINTS3-2: TAAA TATTAGGTACAGAGGCTT; mInts3-1: TTAAAAACAATTTAAAACTCGA; mInts3-2: TACAAATGCAGCTGACAGGAA; mInts3-3: TTCTTATCCTG AAAGGAGGGGA; mInts3-4: TTTAAAACTCGATTATCTTTGC; mInts3-5: TAATCTTACAAGGTCCCGGCCA; mTet3-1: TTATTAAGACCAAACC TGGCTA; mTet3-2: TTAAATGAAGTGTAGGCCATGC; mTet3-3: TT AAATGGAATTTTAAAACTAC; mTet3-4: GCCTGTTAGGCAGATTGTTCT; mTet3-5: GCTCCAACGAGAAGCTATTTG; hUPF1-1: TGGTATTACA GTAAACCACGCA; hUPF1-2: TTGTGATTTAAACTCGTCACCA; mFto-1: TTCTAAGATATAATCCAAGGTG; mFto-2: TCTGGTTTCTGCTGTACTGGTA; mAlkbh5-1: TTGAACTGGAACTTGCAGCCGA; mAlkbh5-2: TTCATCAGCAGCATACCCACTG. mCherry$^+$ or GFP$^+$ cells with shRNAs against hINTS3, mInts3, or mTet3 were FACS-sorted.

**Semi-quantitative and quantitative RT–PCR and mRNA stability assay.** Total RNA was isolated using TRIzol reagent (Life Sciences) with standard RNA extraction protocol for snRNA quantification or using an RNeasy Mini or Micro kit (Qiagen) with DNase I treatment (Qiagen). For cDNA synthesis, total RNA was reverse transcribed with EcoDry kits (Random Hexamer or Oligo dT kits; Clontech), SuperScript (Invitrogen), RNA-Quant cDNA synthesis Kit (System Biosciences), or Verso cDNA Synthesis Kit (Thermo Fisher Scientific). Primers used in reverse-transcriptase polymerase chain reactions (RT–PCR) were: *INTS3* forward1: TGAGTCGTGATGGCATGAAT (exon 4), reverse1: TCTTCACCAGTTCCCGTACC (exon 5; for detection of intron 4 retention), reverse2: CTGCTCTTCAGGACCCACTC (exon 7; for detection of exon 5 skipping); *NDUFAF6* forward: GCCTGTGGCCATTGAACTAT, reverse: ACAATGCCTTGTGCTTTTCC; *PHF21A* forward: TCCATGGCC TGGAACTTTAG, reverse: GCCAGGATGGTGTTCTTCAT; *GLYR1* forward: AGGTCAGGCCCAGTTCTCTT, reverse: TCACGTCTAAGCGTCCAGTG; *GAPDH* forward: GCAAATTCCATGGCACCGTC, reverse: TCGCCCCA CTTGATTTTGG.

The PCR cycling conditions (33 cycles) chosen were as follows: (1) 30 s at 95 °C (2) 30 s at 60 °C (3) 30 s at 72 °C with a final 5-min extension at 72 °C. Reaction products were analysed on 2% agarose gels. The bands were visualized by ethidium bromide staining.

Quantitative real-time reverse transcriptase PCR (qPCR) analyses were performed on an Applied Biosystems QuantStudio 6 Flex cycler using SYBR Green Master Mix (Roche). The following primers were used: *hINTS3*: forward2: CTGCAGGATACCTGCCGTA (exon 4), reverse3: CTTTCCCGTT CCTGACAGAG (intron 5; for specific quantification of transcript with intron 4 retention); forward1: TGAGTCGTGATGGCATGAAT (exon 4), reverse4: GGCTGTAACATCTCCACCTGA (exon 4–6; for specific quantification of transcript with exon 5 skipping); forward3: GGGCAATGCTGAGAGAAG (exon 14), reverse5: TGCCTCTGCATTGTCATAGC (exon 15); *mInts3*: forward: GTGGCGTTATTGACTCTGCAC, reverse: CAGGTTCCCCATCATCACAT; *mFto*: forward: CACTTGGCTTCCTTACCTGACCCCC, reverse: GGTATGCT GCCGGCCTCTCGG; *mAlkbh5*: forward: CGGCCTCAGGACATTAAGGA, reverse: TCGCGGTGCATCTAATCTTG; Total U2snRNA: forward: CTTCTCGG CCTTTTGGCTAAGAT, reverse: GTACTGCAATACCAGGTCGATGC; uncleaved U2snRNA: forward: ACGTCCTCTATCCG+AGGACAATA,

reverse: GCAGGTGCTACCGTCTCTCAC; total U4snRNA: forward: GCAGT ATCGTAGCCAATGAGGTCTA, reverse: CCAGTGCCGACTATATTGCAAGTC.

Uncleaved U4snRNA: forward: CGTAGCCAATGAGGTCTATCCG, reverse: CCTCTGTTGTTCAACTGCAAGAAA; *hGAPDH*: forward: GCAAATT CCATGGCACCGTC, reverse: TCGCCCCACTTGATTTTGG; *mGapdh*: forward: TGGAGAAACCTGCCAAGTATG, reverse: GGAGACAACCTGGTCCTCAG.

All samples, including the template controls were assayed in triplicate. The relative number of target transcripts was normalized to the housekeeping gene found in the same sample. The relative quantification of target gene expression was performed with the standard curve or comparative cycle threshold ($C_T$) method.

mRNA stability assay was performed as previously described[6]. In brief, anti-*UPF1* shRNA- or control shRNA lentivirus-infected K562 *SRSF2^{P95H}* knock-in cells were generated by puromycin selection (1 μg ml$^{-1}$) for 7 days and shRNAs against *UPF1* were expressed by doxycycline (2 μg ml$^{-1}$) for 2 days. GFP (shRNA)-positive cells were FACS-sorted, treated with 2.5 μg ml$^{-1}$ actinomycin D (Life Technologies), and collected at 0, 2, 4, 8, and 12 h.

**ChIP assays.** Cells were crosslinked and collected. Chromatin was broken down into 200–1,000-bp fragments using an E220 Focused-ultrasonicator. An antibody was added into the lysate and incubated overnight at 4 °C. Twenty microlitres of ChIP–grade Protein A/G Dynabeads was added into each IP tube and incubated for 2 h. IP samples were washed and crosslinks reversed by adding proteinase K and incubating overnight at 65 °C. DNA was purified with AMPureXP beads and eluted DNA was subjected to qPCR to measure the enrichment. RNAPII antibody (05-623; EMD Millipore) was used in this study. Primer sequences used for ChIP-PCR were as follows: Intron 3-1 forward: atacccggcccttgctatac, reverse: gcaacttccttagcctgctg; Intron 3-1 forward: atacccggcccttgctatac, reverse: gcaacttccttagcctgctg; Intron 3-2 forward: ctggcaggtgaaaagcagat, reverse: ggcaggggagagaaaagc; Intron 3-3 forward: agcaggcttttctgcctcat, reverse: tttcttttccacagggtcct; Exon 4 forward: cgggacttagctctggtgag, reverse: cctgagtacggcaggtatcc; Intron 4 forward: ctct gtcaggaacgggaaag, reverse: tgtgagtttgagaagggagcta; Exon 5 forward: acgggaactggtgaa gagtg, reverse: ctgggctctcctccttttctt; Intron 5-1 forward: ctccacccccattatctgaa, reverse: aaatgtcagggtcgttctgtg; Intron 5-2 forward: tcggtgacatctgtctgagc, reverse: cagtgggctaa tggtgaggt; Intron 5-3 forward: aacactgatgctcctgttttga, reverse: actatgccttgccccaggt; Intron 5-4 forward: gctgttgtcagccacctgta, reverse: tttggcccttgaaaatgaac; Intron 5-5 forward: tgtgttaattctgccccaca, reverse: ggatgtcctgagtcctgcac; Intron 5-6 forward: gtaatgggatggcagtcagg, reverse: cctgatttcaaaaggggaaa; Exon 6 forward: agcaaaggtagc atccacca, reverse: cttgcctccccctctctaac; Intron 6-1 forward: tttgatccagacctccttgg, reverse: gcaggggagaaaaggatacc; Intron 6-2 forward: gggggtacatattgggcttt, reverse: gaaagcctcacctccaaaca; Intron 6-3-CTCF binding site forward: ctcctcccaacgttcacact, reverse: atccgtgcccagagcacta; Intron 6-4 forward: agggggcctttcaactctt, reverse: atggggacaggacgtatttg; Intron 6-5 forward: ttccctgccttccaacag, reverse: tcccagttgctt taaaaggagt.

ChIP–seq libraries were prepared as previously described[31] and sequenced by the Integrated Genomics Operation (IGO) at MSK with 50 bp paired-end reads.

**ChIP–seq of primary human AML samples.** ChIP was performed as previously described[32] using the following antibodies: RNAPolII-Ser2P antibody ChIP Grade (Abcam ab5095), RNAPlI-Ser5P antibody [4H8] (Abcam ab5408), and anti-HP1γ antibody, clone 42s2 (05-690 from Merck Millipore). Libraries were size-selected with AMPure beads (Beckman Coulter) for 200–800-bp size range and quantified by qPCR using a KAPA Library Quantification Kit. ChIP–seq data were generated using the NextSeq platform from Illumina with 2 × 75 bp Hi Output (all samples pooled, and sequenced on four consecutive runs before merger of FASTQ files).

**Histological analyses.** Mice were euthanized and autopsied, and dissected tissue samples were fixed in 4% paraformaldehyde, dehydrated, and embedded in paraffin. Paraffin blocks were sectioned at 4 μm and stained with haematoxylin and eosin (H&E). Images were acquired using an Axio Observer A1 microscope (Carl Zeiss) or scanned using a MIRAX Scanner (Zeiss).

**Patient samples.** Studies were approved by the Institutional Review Boards of Memorial Sloan Kettering Cancer Center (under MSK IRB protocol 06-107), Université Paris-Saclay (under declaration DC-200-725 and authorization AC-2013-1884), and the University of Manchester (institution project approval 12-TISO-04), and conducted in accordance with the Declaration of Helsinki protocol. Written informed consent was obtained from all participants. Manchester samples were retrieved from the Manchester Cancer Research Centre Haematological Malignancy Tissue Biobank, which receives sample donations from all consenting patients with leukaemia presenting to The Christie Hospital (REC Reference 07/H1003/161+5; HTA license 30004; instituted with approval of the South Manchester Research Ethics Committee). Patient samples were anonymized by the Hematologic Oncology Tissue Bank of MSK, Biobank of Gustave Roussy, and the Manchester Cancer Research Centre Haematological Malignancy Tissue Biobank.

**Mutational analysis of patient samples.** Genomic DNA is routinely extracted from mononuclear cell samples submitted to the Manchester Cancer Research Centre Haematological Tissue Biobank. Targeted sequencing for recurrent myeloid mutations, using either: (a) a 54 gene panel (TruSight Myeloid; Illumina),

pooling 96 samples with 5% PhiX onto a single NextSeq high output, $2 \times 151$-bp sequencing run; variant call format (VCF) files were analysed using Illumina's Variant Studio software; (b) a 40 gene panel (Oncomine Myeloid Research Assay; ThermoFisher), processing 8 samples per Ion 530 chip on the IonTorrent platform; data analysis performed using the Ion Reporter software; (c) a 27 gene custom panel ($48 \times 48$ Access Array; Fluidigm) sequenced by Leeds HMDS on the MiSeq platform (300v2); or (d) MSK HemePACT[33] targeting all coding regions of 585 genes known to be recurrently mutated in leukaemias, lymphomas, and solid tumours. All panels provide sufficient coverage to detect minimum variant allele fraction 5% for all genes, except for the Access Array panel and *SRSF2*; all samples genotyped by this approach underwent manual Sanger sequencing of *SRSF2* exon 1 using the following primers (tagged with Fluidigm Access Array sequencing adaptors CS1/CS2): forward: acactgacgacatggttctacacccgtttacctgcggctc, reverse: tacggtagc agagacttggtctccttcgttcgctttcacgacaa.

**Statistics and reproducibility.** Statistical significance was determined by (1) unpaired two-sided Student's *t*-test after testing for normal distribution, (2) one-way or two-way ANOVA followed by Tukey's, Sidak's, or Dunnett's multiple comparison test, or (3) Kruskal–Wallis tests with uncorrected Dunn's test where multiple comparisons should be adjusted (unless otherwise indicated). Data were plotted using GraphPad Prism 7 software as mean values, with error bars representing standard deviation. For categorical variables, statistical analysis was done using Fisher's exact test or $\chi^2$-test (two-sided). Representative western blot and PCR results are shown from three or more than three biologically independent experiments. Representative flow cytometry results and cytomorphology are shown from biological replicates ($n \geq 3$). $*P < 0.05$, $**P < 0.01$ and $***P < 0.001$, respectively, unless otherwise specified.

**mRNA isolation, sequencing, and analysis.** RNA was extracted as shown above. Poly(A)-selected, unstranded Illumina libraries were prepared with a modified TruSeq protocol. $0.5\times$ AMPure XP beads were added to the sample library to select for fragments <400 bp, followed by $1\times$ beads to select for fragments >100 bp. These fragments were then amplified with PCR (15 cycles) and separated by gel electrophoresis (2% agarose). DNA fragments 300 bp in length were isolated and sequenced on an Illumina HiSeq 2000 (about $100 \times 10^6$ 101-bp reads per sample).

Primary samples from the Manchester Cancer Research Centre Haematological Malignancies Biobank with known *IDH2/SRSF2* mutation genotype were FACS-sorted to enrich for blasts on a FACS Aria III sorter using a panel including the following antibodies (all mouse anti-human): CD34-PerCP (8G12, BD); CD117-PECy7 (104D2, eBioscience); CD33-APC (P67.6, BioLegend); HLA-DR-FITC (L243, BioLegend); CD13-PE (L138, BD); CD45-APC-H7 (2D1, BD). RNA was extracted immediately using a Qiagen Micro RNeasy kit. All RNA samples had RIN values > 8. Poly(A)-selected, strand-specific SureSelect (Agilent) mRNA libraries were prepared using 200 ng RNA according to the manufacturer's protocol. Libraries were pooled and sequenced ($2 \times 101$ bp paired end) to >100 million reads per sample on two HiSeq 2500 high throughput runs before retrospective merger of FASTQ files for downstream alignment and splicing analysis as described below. Transcriptional analysis was done using gene set enrichment analysis (GSEA)[34].

**Publicly available RNA-seq data.** Unprocessed RNA-seq reads of TCGA and Leucegene datasets (patients with AML) were downloaded from NCI's Genomic Data Commons Data Portal (GDC Legacy Archive; TCGA-LAML dataset) and NCBI's Sequence Read Archive (SRA; accession number SRP056295). The TCGA dataset consists of paired-end $2 \times 50$-bp libraries, with an average read count of 76.92 M. The Leucegene dataset consists of paired-end $2 \times 100$-bp libraries, with an average read count of 50.40 M per sample. The RNA-seq samples in the Leucegene dataset have 1–3 sequencing runs (about 50 M each run), and only one run was used to represent each RNA-seq sample.

**Genome and splice junction annotations.** Human assembly hg38 (GRCh38) and Ensembl database (human release 87) were used as the reference genome and gene annotation, respectively. RNA-seq reads were aligned by using 2-pass STAR 2.5.2a[35]. Known splice junctions from the gene annotation and new junctions identified from the alignments of the TCGA dataset were combined to create the database of alternative splicing events for splicing analysis.

**Mutational analysis for the RNA-seq data.** Samtools (1.3.1) was used to generate VCF files for seven target genes: *IDH1*, *IDH2*, *TET2*, *SF3B1*, *SRSF2*, *U2AF1*, and *ZRSR2* with mpileup parameters (-Bvu). The VCF files were further processed by our in-house scripts to filter out mutations with a variant allele frequency (VAF) lower than 15%. The filtered VCF files were used for variant effect predictor (v.89.4) to annotate the consequences of the mutations. We defined control patient samples as those without mutations in the seven target genes, *IDH2* mutated samples as those with only *IDH2* mutations but no mutations in the other six target genes, *SRSF2* mutated samples as those with only *SRSF2* mutations but no mutations in the other six target genes, Double-mutant samples as those with both *IDH2* and *SRSF2* mutations but no mutations in the other five target genes, and 'others' as those with mutations in *IDH1*, *TET2*, *SF3B1*, *U2AF1*, and *ZRSR2*.

**Identification and quantification of differential splicing.** The inclusion ratios of alternative exons or introns were estimated by using PSI-Sigma[25]. In brief, the new PSI index considers all isoforms in a specific gene region and can report the PSI value of individual exons in a multiple-exon-skipping or more complex splicing event. The database of splicing events was constructed based on both gene annotation and the alignments of RNA-seq reads. A new splicing event not known to the gene annotation is labelled as 'novel' and a splicing event with a reference transcript that is known to induce nonsense-mediated decay is labelled as 'NMD' in Supplementary Tables. The inclusion ratio of an intron retention isoform is estimated based on the median of 5 counts of intronic reads at the 1st, 25th, 50th, 75th and 99th percentiles in the intron. A splicing event is reported when both sample-size and statistical criteria are satisfied. The sample-size criterion requires a splicing event to have more than 20 supporting reads in more than 75% of the 2 populations in the comparison. For example, for a comparison of 130 control versus 6 *IDH2* mutant samples, a splicing event would be reported only when having more than 98 controls and 5 *IDH2* mutant samples with more than 20 supporting reads. In addition, a splicing event is reported only when it has more than 10% PSI change in the comparison and has a *P* value lower than 0.01.

To generate Fig. 4f, RNA-seq reads were mapped and PSI values were calculated using junction-spanning reads as previously described[36,37]. All reads mapping to the INTS3 introns (chr1:153,718,433-153,722,231; hg19) were extracted from the .bam files and the per-nucleotide coverage was calculated. Data from normal peripheral blood and BM mononuclear cells and CD34$^+$ cord blood cells are combined and shown as normal haematopoietic cells.

**Motif enrichment and distribution.** Motif analysis was done by using MEME SUITE[38]. In brief, the sequences of alternative exons of exon-skipping events were extracted from a given strand of the reference genome. The sequences were used as the input for MEME SUITE to search for motifs. One occurrence per sequence was set to be the expected site distribution. The width of motif was set to 5. The top 1 motif was selected on the basis of the ranking of *E*-value.

**Heat map and sample clustering (differential splicing).** The heat maps and sample clustering were done by using MORPHEUS (https://software.broadinstitute.org/morpheus/). The individual values in the matrix for the analysis were PSI values of a splicing event from a given RNA-seq sample. Splicing events were selected based on three criteria: (1) present in both TCGA and Leucegene datasets; (2) more than 15% PSI changes; and (3) false discovery rate smaller than 0.01. Unsupervised hierarchical clustering was based on one minus Pearson's correlation (complete linkage).

**Correlation between global changes in splicing and DNA methylation.** DNA methylation levels were determined by eRRBS and differentially spliced events were obtained from RNA-seq data. In Fig. 3e, Overlaps of differentially methylated regions of DNA with differential splicing was obtained by evaluating differential cytosine methylation in 500-bp segments of DNA at genomic coordinates at which differential RNA splicing were observed comparing AML with distinct *IDH2/SRSF2* genotypes shown (WT represents patients without mutations in *IDH1/IDH2*/spliceosomal genes).

**Reporting summary.** Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

## Data availability

RNA-seq, ChIP–seq and eRRBS data have been deposited in the NCBI Sequence Read Archive under accession number SRP133673. Gel source data are shown in Supplementary Fig. 1. Other data that support the findings of this study are available from the authors upon reasonable request.

25. Lin, K. T. & Krainer, A. R. PSI-Sigma: a comprehensive splicing-detection method for short-read and long-read RNA-seq analysis. *Bioinformatics* btz438 (2019).
26. Moran-Crusio, K. et al. Tet2 loss leads to increased hematopoietic stem cell self-renewal and myeloid transformation. *Cancer Cell* **20**, 11–24 (2011).
27. Shih, A. H. et al. Combination targeted therapy to disrupt aberrant oncogenic signaling and reverse epigenetic dysfunction in *IDH2*- and *TET2*-mutant acute myeloid leukemia. *Cancer Discov.* **7**, 494–505 (2017).
28. Georgiades, P. et al. *VavCre* transgenic mice: a tool for mutagenesis in hematopoietic and endothelial lineages. *Genesis* **34**, 251–256 (2002).
29. Zuber, J. et al. Toolkit for evaluating genes required for proliferation and survival using tetracycline-regulated RNAi. *Nat. Biotechnol.* **29**, 79–83 (2011).
30. Lee, M. et al. Engineered split-TET2 enzyme for inducible epigenetic remodeling. *J. Am. Chem. Soc.* **139**, 4659–4662 (2017).
31. Kleppe, M. et al. Dual targeting of oncogenic activation and inflammatory signaling increases therapeutic efficacy in myeloproliferative neoplasms. *Cancer Cell* **33**, 29–43 (2018).
32. Maiques-Diaz, A. et al. Enhancer activation by pharmacologic displacement of LSD1 from GFI1 induces differentiation in acute myeloid leukemia. *Cell Rep.* **22**, 3641–3659 (2018).
33. Cheng, D. T. et al. Memorial Sloan Kettering–integrated mutation profiling of actionable cancer targets (MSK–IMPACT): a hybridization capture-based next-generation sequencing clinical assay for solid tumor molecular oncology. *J. Mol. Diagn.* **17**, 251–264 (2015).

34. Subramanian, A. et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl Acad. Sci. USA* **102**, 15545–15550 (2005).
35. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
36. Dvinge, H. & Bradley, R. K. Widespread intron retention diversifies most cancer transcriptomes. *Genome Med.* **7**, 45 (2015).
37. Hubert, C. G. et al. Genome-wide RNAi screens in human brain tumor isolates reveal a novel viability requirement for PHF5A. *Genes Dev.* **27**, 1032–1045 (2013).
38. Bailey, T. L. et al. MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res.* **37**, W202–W208 (2009).
39. Robinson, J. T. et al. Integrative genomics viewer. *Nat. Biotechnol.* **29**, 24–26 (2011).
40. Intlekofer, A. M. et al. Hypoxia induces production of l-2-hydroxyglutarate. *Cell Metab.* **22**, 304–311 (2015).
41. Dvinge, H. et al. Sample processing obscures cancer-specific alterations in leukemic transcriptomes. *Proc. Natl Acad. Sci. USA* **111**, 16802–16807 (2014).
42. Macrae, T. et al. RNA-seq reveals spliceosome and proteasome genes as most consistent transcripts in human cancer cells. *PLoS ONE* **8**, e72884 (2013).

**Competing interests** A.M.I. has served as a consultant and advisory board member for Foundation Medicine. E.M.S. has served on advisory boards for Astellas Pharma, Daiichi Sankyo, Bayer, Novartis, Syros, Pfizer, PTC Therapeutics, AbbVie, Agios and Celgene and has received research support from Agios, Celgene, Syros and Bayer. R.L.L. is on the Supervisory Board of Qiagen and the Scientific Advisory Board of Loxo, reports receiving commercial research grants from Celgene, Roche and Prelude, has received honoraria from the speakers bureaus of Gilead and Lilly, has ownership interest (including stock and patents) in Qiagen and Loxo, and is a consultant and/or advisory board member for Novartis, Roche, Janssen, Celgene and Incyte. A.R.K. is a founder, director, advisor, stockholder and chair of the Scientific Advisory Board of Stoke Therapeutics and receives compensation from the company. A.R.K. is a paid consultant for Biogen; he is a member of the SABs of Skyhawk Therapeutics, Envisagenics BioAnalytics and Autoimmunity Biologic Solutions, and has received compensation from these companies in the form of stock. A.R.K. is a research collaborator of Ionis Pharmaceuticals and has received royalty income from Ionis through his employer, Cold Spring Harbor Laboratory. O.A.-W. has served as a consultant for H3 Biomedicine, Foundation Medicine, Merck and Janssen. O.A.-W. has received personal speaking fees from Daiichi Sankyo. O.A.-W. has received previous research funding from H3 Biomedicine unrelated to the current manuscript. D.I., R.K.B. and O.A.-W. are inventors on a provisional patent application (patent number FHCC.P0044US.P) applied for by Fred Hutchinson Cancer Research Center on the role of reactivating BRD9 expression in cancer by modulating aberrant *BRD9* splicing in *SF3B1* mutant cells.

**Additional information**
**Supplementary information** is available for this paper at https://doi.org/10.1038/s41586-019-1618-0.
**Correspondence and requests for materials** should be addressed to O.A.-W.
**Peer review information** *Nature* thanks Rotem Karni and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.
**Reprints and permissions information** is available at http://www.nature.com/reprints.
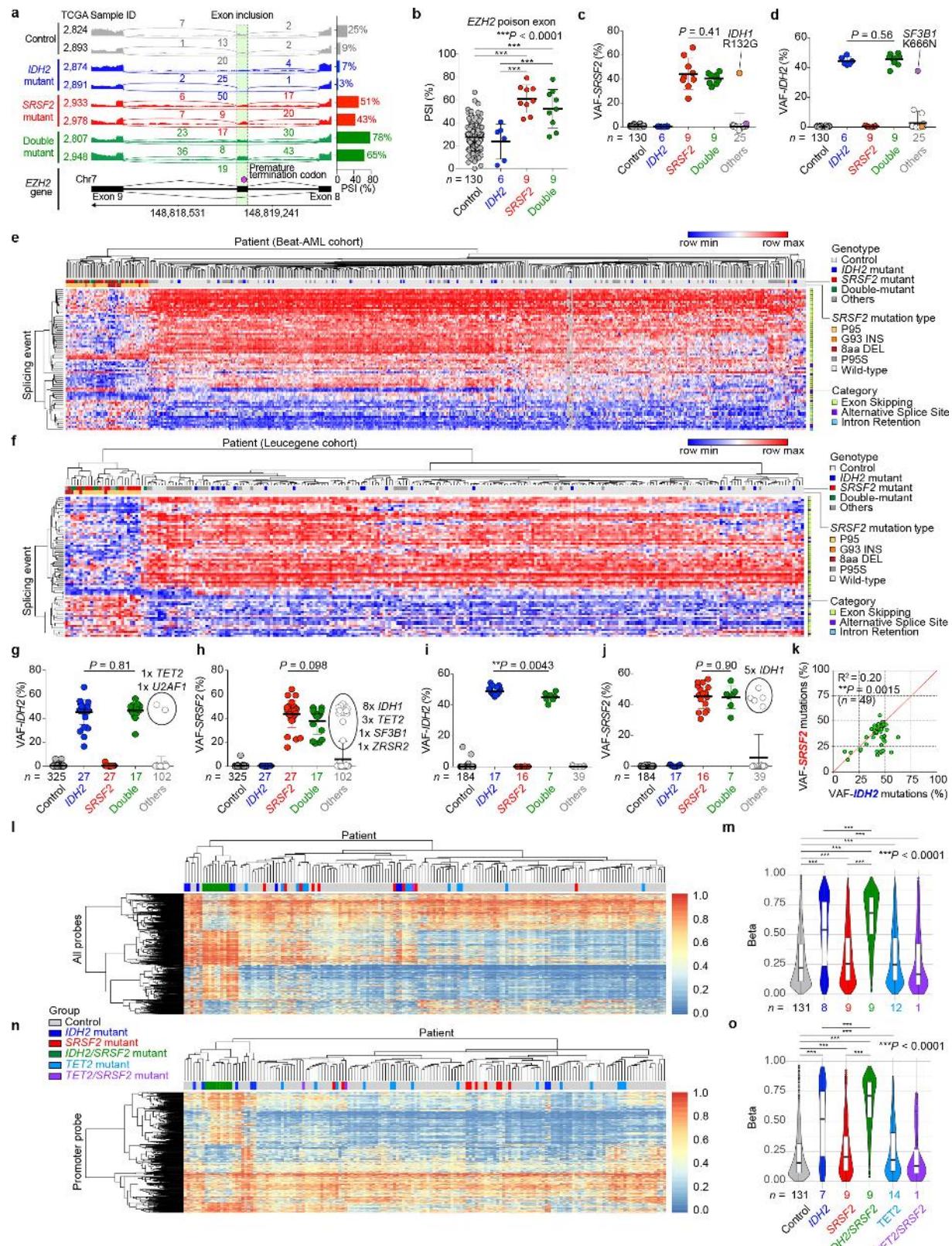
**Extended Data Fig. 1 |** See next page for caption.

**Extended Data Fig. 1 | Mutant *SRSF2*-mediated splicing events in acute myeloid leukaemia (AML). a**, Representative Sashimi plots of RNA-seq data from the TCGA showing the poison exon inclusion event in *EZH2* ('Control' represents samples that are wild type (WT) for the following seven genes: *IDH1*, *IDH2*, *TET2*, *SRSF2*, *SF3B1*, *U2AF1*, and *ZRSR2*; '*IDH2* mutant' refers to patients with an *IDH2* mutation and no mutation in the other six genes; '*SRSF2* mutant' refers to patients with an *SRSF2* mutation and no mutation in the other six genes; 'double-mutant' refers to patients with an *IDH2* and *SRSF2* mutation and no mutation in the other five genes; 'others' refers to patients with mutations in *IDH1*, *TET2*, *SF3B1*, *U2AF1* or *ZRSR2*; figure made using Integrative Genomics Viewer (IGV 2.3)[39]). **b**, PSI values of *EZH2* poison exon inclusion (the number of analysed patients is indicated; mean ± s.d.; one-way ANOVA with Tukey's multiple comparison test). Note that patients classified as 'others' include one patient with an *SRSF2*[P95L] mutation with a coexisting *IDH1*[R132G] mutation (TCGA ID: 2990) and one patient with an *IDH2*[R140Q] mutation also having an *SF3B1*[K666N] mutation (TCGA ID: 2973), which were excluded from the analyses shown above. **c, d, g–j**, VAFs of *SRSF2* mutations affecting the proline 95 residue (**c, h, j**) and *IDH2* mutations affecting *IDH2* arginine 140 or 172 (**d, g, i**) in TCGA (**c, d**), Beat AML (**g, h**) and Leucegene (**i, j**) datasets (mean ± s.d.; two-sided Student's *t*-test). **e, f**, Heat map based on the ΔPSI of mutant *SRSF2*-specific splicing events in AML from Beat AML (**e**) and Leucegene (**f**) cohorts. '8aa DEL' represents samples with 8 amino acid deletions in SRSF2 starting from proline 95, which has similar effects on splicing as point mutations affecting SRSF2 P95. Detailed information of splicing events shown is available in Supplementary Table 1. **k**, VAFs of *IDH2* (*x* axis) and *SRSF2* mutations (*y* axis) in *IDH2* and *SRSF2* double-mutant AML determined by RNA-seq data from the TCGA, Beat AML, Leucegene and our previously unpublished cohorts (Pearson correlation coefficient; *P* value (two-tailed) was calculated by Prism7). **l, n**, Unsupervised hierarchical clustering of DNA methylation levels of all probes (**l**) or at the promoter probes (**n**) in the TCGA AML cohort based on *IDH2*, *SRSF2* and *TET2* genotypes. **m, o**, DNA methylation levels of AML samples from each genotype are quantified and visualized from **l** and **n** as violin plots (the line represents mean, box edges show 25th and 75th percentiles and whiskers represent 2.5th and 97.5th percentiles; one-way ANOVA with Tukey's multiple comparison test). \*\**P* < 0.01, \*\*\**P* < 0.001.
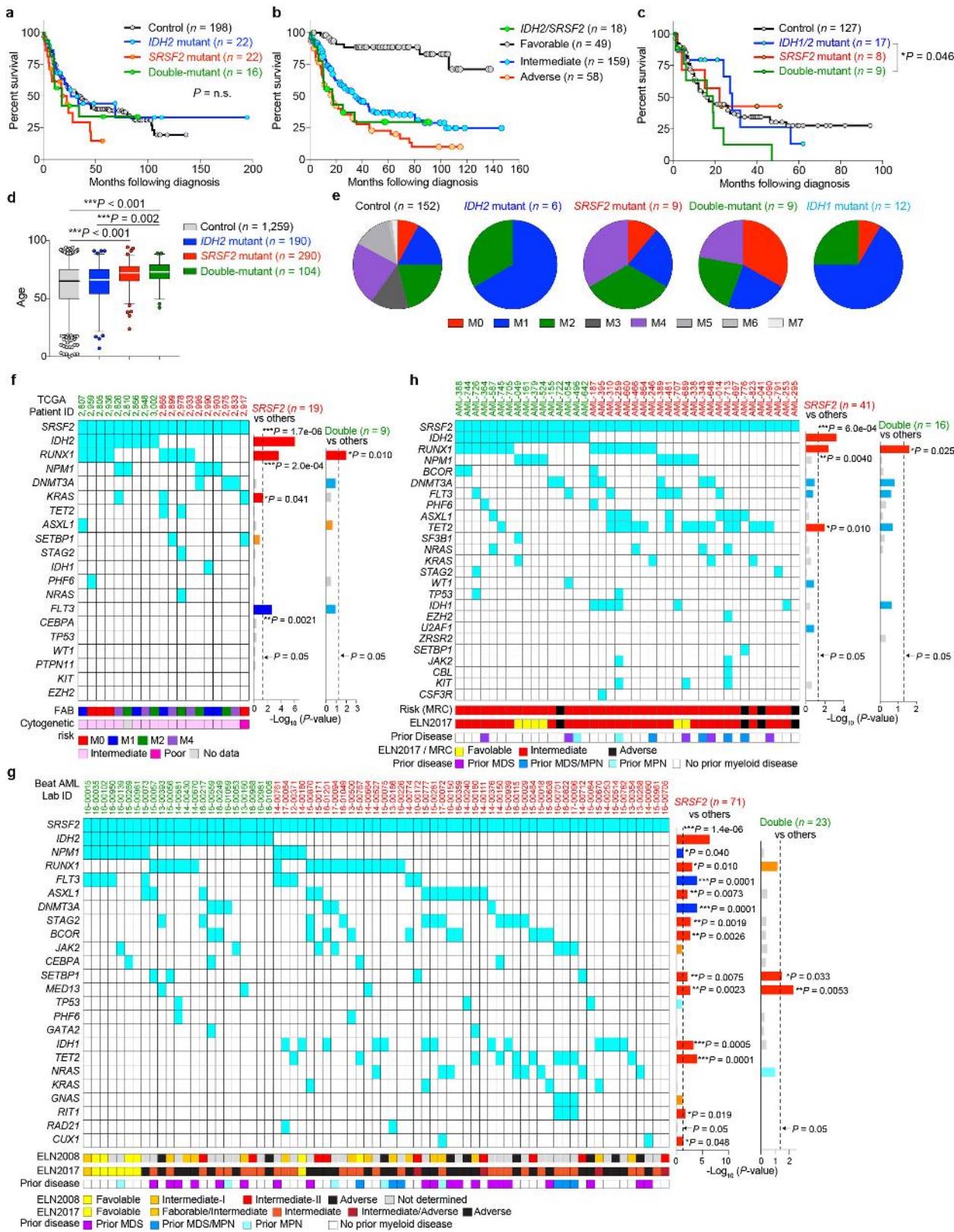
**Extended Data Fig. 2 |** See next page for caption.

**Extended Data Fig. 2 | Clinical relevance of coexisting *IDH2* and *SRSF2* mutations in AML. a–c**, Kaplan–Meier survival analysis of patients with AML from the Manchester/Christie Biobank dataset (**a**: based on *IDH2* and *SRSF2* genotype ($n = 258$); **b**: based on cytogenetic risk ($n = 284$)) and the TCGA (**c**; $n = 161$; based on *IDH1*, *IDH2* and *SRSF2* genotypes (log-rank (Mantel–Cox) test (two-sided)). **d**, Age at diagnosis of patients from the TCGA, Beat AML, and Manchester/Christie Biobank cohorts combined (the line represents mean, box edges show 25th and 75th percentiles and whiskers represent 2.5th and 97.5th percentiles; samples below 2.5th percentile and above 97.5th percentile are shown as dots; one-way ANOVA with Tukey's multiple comparison test). **e**, Distribution of French–American–British (FAB) classification of patients with AML with the indicated genotypes from the TCGA cohort. **f–h**, Mutations coexisting with *IDH2* and *SRSF2* double-mutant and *SRSF2* single-mutant AML from the TCGA (**f**), Beat AML (**g**), and Manchester/Christie Biobank (**h**) cohorts are shown with FAB classification, cytogenetic risk, prior history of myeloid disorders, and genetic risk stratification based on European LeukaemiaNet (ELN) 2008 and ELN2017 guidelines (the number of patients is indicated; *P* values on the right represent statistical significance of co-occurrence (red and orange) or mutual exclusivity (blue and light blue) of each gene mutation with *SRSF2* (including those in *IDH2* and *SRSF2* double-mutant AML) or coexisting *IDH2* and *SRSF2* mutations; Fisher's exact test (two-sided)). $*P < 0.05$, $**P < 0.01$, $***P < 0.001$.
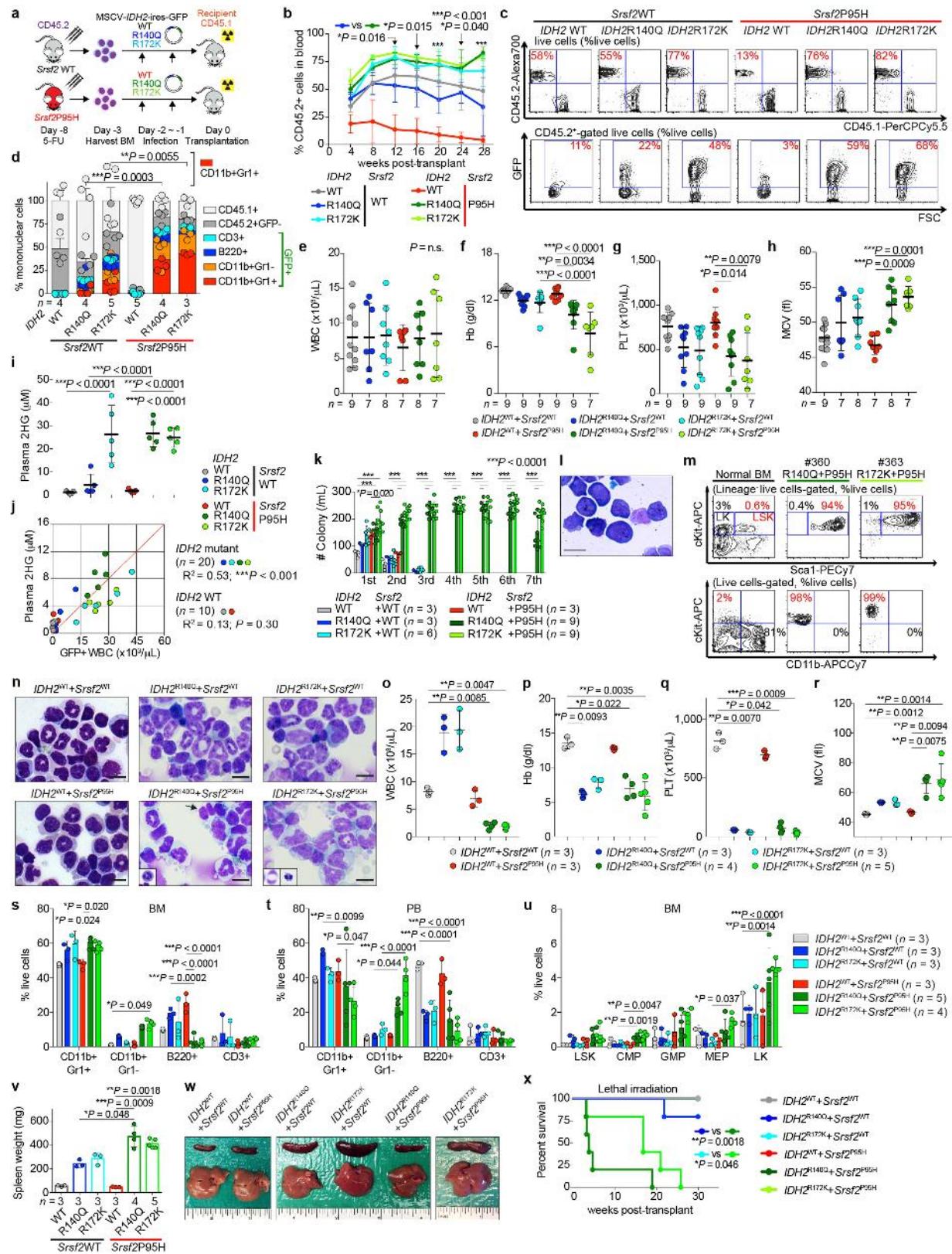
**Extended Data Fig. 3 |** See next page for caption.

**Extended Data Fig. 3 | Mutant *IDH2* cooperates with mutant *Srsf2* to generate lethal MDS with proliferative features in vivo. a**, Schematic of BM transplantation model. **b, c**, Chimerism of CD45.2$^+$ cells in the peripheral blood of recipient mice over time (**b**) ($n = 5$ per group at 4 weeks; mean percentage $\pm$ s.d.; two-way ANOVA with Tukey's multiple comparison test) and representative flow cytometry data showing the chimerism of CD45.2$^+$ versus CD45.1$^+$ (top) or GFP$^+$ (bottom) cells in peripheral blood at 16 weeks post-transplant (**c**) (representative results from five recipient mice; the percentages listed represent the percent of cells within live cells). **d**, Composition of peripheral blood mononuclear cells (PBMNCs) at 28 weeks post-transplant (the number of analysed mice is indicated; mean + s.d.; two-way ANOVA with Tukey's multiple comparison tests; statistical significances were detected in percentage of CD11b$^+$Gr1$^+$ cells in *IDH2$^{R140Q}$* + *Srsf2$^{WT}$* versus *IDH2$^{R140Q}$* + *Srsf2$^{P95H}$* and in *IDH2$^{R172K}$* + *Srsf2$^{WT}$* versus *IDH2$^{R172K}$* + *Srsf2$^{P95H}$*). **e–h**, Blood counts at 20 weeks post-transplant (white blood cells (WBC) (**e**); haemoglobin (Hb) (**f**); platelets (PLT) (**g**); mean corpuscular volume (MCV) (**h**); the number of analysed mice is indicated; mean $\pm$ s.d.; one-way ANOVA with Tukey's multiple comparison tests). **i**, Plasma 2HG levels at 20 weeks post-transplant (2HG levels were quantified as previously described[40]; $n = 5$ per group were randomly selected; mean $\pm$ s.d.; one-way ANOVA with Tukey's multiple comparison test). **j**, Correlations between plasma 2HG levels and number of GFP$^+$ cells in peripheral blood at 24 weeks post-transplant ($n = 5$ per group; the Pearson correlation coefficient ($R^2$) and *P* values (two-tailed) were calculated using PRISM 7). **k**, Colony numbers from serial replating assays of BM cells collected from end-stage mice from Fig. 2b are shown (mean value $\pm$ s.d. represented by lines above the box; the number of analysed mice is indicated; two-way ANOVA with Tukey's multiple comparison test). **l**, Giemsa staining of *IDH2$^{R140Q}$ Srsf2$^{P95H}$* double-mutant cells from the sixth plating (scale bar, 10 μm; original magnification, ×400;

representative result from 9 biologically independent experiments). **m**, Immunophenotype of colony cells at the sixth plating. Normal BM cells were used as a control (the percentage listed represent the percent of cells within live cells; representative result from nine recipient mice). **n**, Cytomorphology of BM mononuclear cells (BMMNCs) from recipient mice at end stage. BM cells from *IDH2* single-mutant and *IDH2* and *Srsf2* double-mutant groups have increased granulocytes. In addition, *IDH2* and *Srsf2* double-mutant groups had proliferation of monoblastic and monocytic cells as well as dysplastic features such as abnormally segmented neutrophils (black arrow and inset) and binucleated erythroid precursors with irregular nuclear contours (insets) (scale bar, 10 μm; original magnification, ×400; representative results from 3 controls and 9 recipients are shown; number of mice indicated in **o–r**). **o–r**, Blood counts at end-stage (WBC (**o**); Hb (**p**); PLT (**q**); MCV (**r**); the number of analysed mice is indicated; mean $\pm$ s.d.; Kruskal–Wallis tests with uncorrected Dunn's test). **s–u**, Results from flow cytometry analysis of BM (**s**) and peripheral blood (**t**) mature lineages as well as BM haematopoietic stem/progenitor cells (HSPC) from two tibias, two femurs, and two pelvic bones (**u**) are quantified (LSK, Lineage$^-$SCA1$^+$KIT$^+$; LT-HSC, long-term haematopoietic stem cell (HSC); ST-HSC, short-term HSC; MPP, multi-potent progenitor; LK: Lineage$^-$SCA1$^-$KIT$^+$; CMP, common myeloid progenitor; GMP, granulocyte-monocyte progenitor; MEP, megakaryocyte-erythroid progenitor; the number of analysed mice is indicated; mean + s.d.; two-way ANOVA with Tukey's multiple comparison test). **v, w**, Spleen weight at end stage (**v**; the number of analysed mice is indicated; mean $\pm$ s.d.; two-way ANOVA with Tukey's multiple comparison test) and representative photographs of spleens from recipient mice from **v** (**w**; each photograph was taken with an inch ruler). **x**, Kaplan–Meier survival analysis of serially transplanted recipient mice that were lethally irradiated ($n = 5$ per group; log-rank (Mantel–Cox) test (two-sided). *$P < 0.05$, **$P < 0.01$, ***$P < 0.001$.
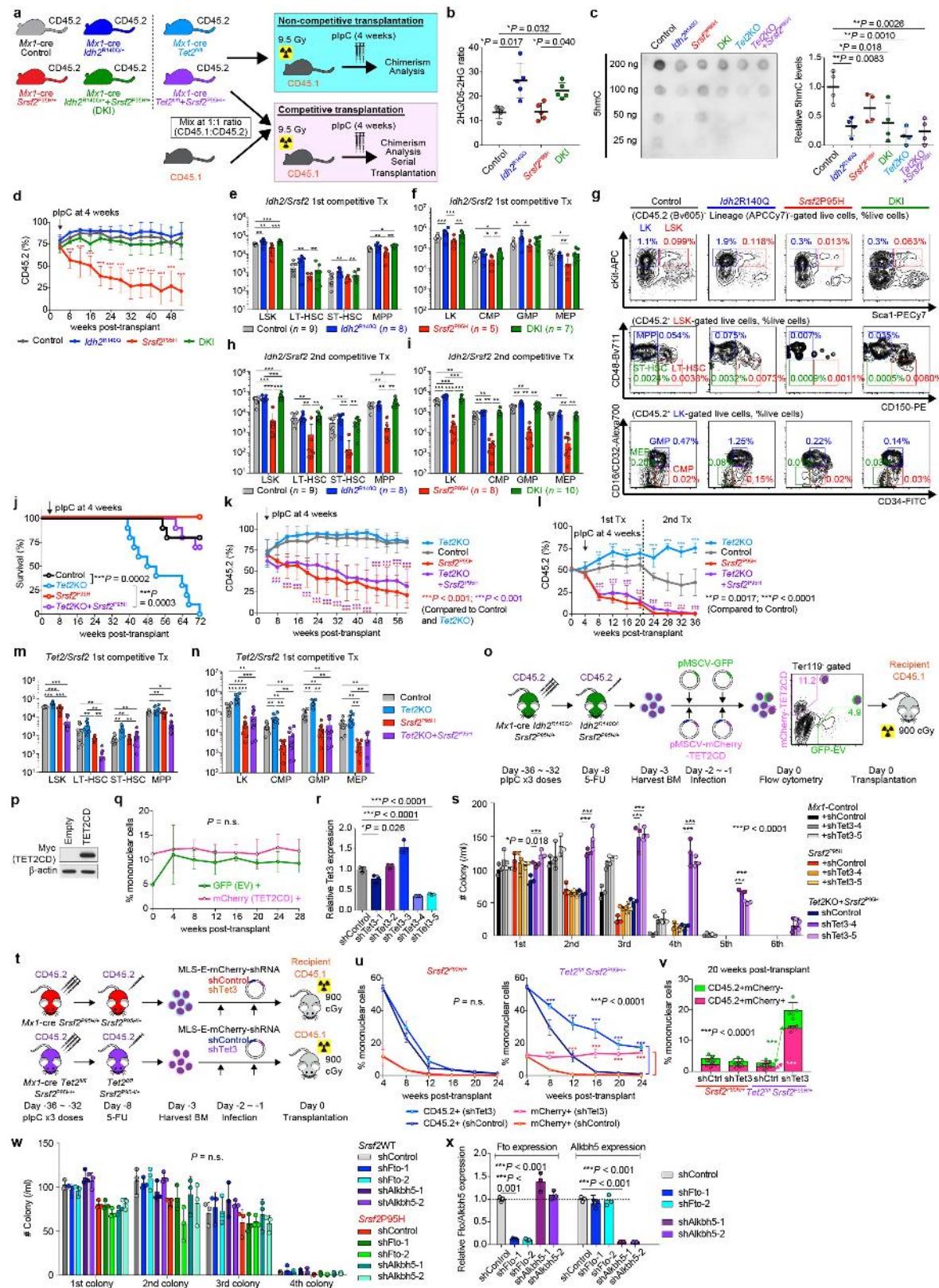
**Extended Data Fig. 4** | See next page for caption.

**Extended Data Fig. 4 | Collaborative effects of mutant *Idh2* and mutant *Srsf2* are not dependent on *Tet2* loss alone. a**, Schematic of competitive and non-competitive transplantation assays of CD45.2$^+$ *Mx1-cre* control, *Mx1-cre Idh2$^{R140Q/+}$*, *Mx1-cre Srsf2$^{P95H/+}$*, *Mx1-cre Idh2$^{R140Q/+}$Srsf2$^{P95H/+}$* mice, *Mx1-cre Tet2$^{fl/fl}$*, *Mx1-cre Tet2$^{fl/fl}$Srsf2$^{P95H/+}$* mice into CD45.1$^+$ recipient mice. **b**, 2HG levels of bulk PBMNCs from primary *Mx1-cre* mice were measured at three months post-pIpC (polyinosinic:polycytidylic acid) and normalized to internal standard (D-2-hydroxyglutaric-2,3,3,4,4-*d5* acid; D5-2HG) (2HG and D5-2HG levels were quantified as described[40]; $n = 5$ per group; mean ± s.d.; one-way ANOVA with Tukey's multiple comparison test). **c**, DNA extracted from sorted KIT$^+$ BM cells from primary *Mx1-cre* mice at one month post-pIpC was probed with antibodies specific for 5-hydroxymethylcytosine (5hmC) (left). Relative intensity of each dot was measured by ImageJ and divided by input DNA amount for comparison (right; $n = 4$; intensity of each dot divided by amount of input DNA was combined per genotype; representative results from 3 biologically independent experiments with similar results; mean ± s.d.; one-way ANOVA with Tukey's multiple comparison test). **d**, Chimerism of peripheral blood CD45.2$^+$ cells in non-competitive transplantation (pIpC was injected at 4 weeks post-transplant; mean ± s.d.; $n = 10$ (control and *Idh2$^{R140Q}$*), $n = 8$ (*Srsf2$^{P95H}$*), and $n = 9$ (DKI) at 0 week; two-way ANOVA with Tukey's multiple comparison test; *P* values from comparison between *Srsf2$^{P95H}$* and each of other groups are shown). **e–i**, Absolute number of BM HSPCs from two tibias, two femurs, and two pelvic bones were measured in the primary (**e**, **f**) and serial (**h**, **i**) competitive transplant of *Idh2* and *Srs2* mutant cells, and representative flow cytometry of BM HSPCs from the primary competitive transplant of *Idh2* and *Srsf2* mutant cells from **e**, **f** (the percentage listed represents the percent of cells within live cells) (the number of analysed mice is indicated; mean + s.d.; two-way ANOVA with Tukey's multiple comparison test). **j**, Kaplan–Meier survival analysis of CD45.1$^+$ recipient mice transplanted non-competitively with BM cells from CD45.2$^+$ *Mx1-cre* control, *Mx1-cre Tet2$^{fl/fl}$*, *Mx1-cre Srsf2$^{P95H/+}$*, and *Mx1-cre Tet2$^{fl/fl}$Srsf2$^{P95H/+}$* mice (pIpC was injected at 4 weeks post-transplant; $n = 10$ per genotype; log-rank (Mantel–Cox) test (two-sided)). **k**, **l**, Chimerism of peripheral blood CD45.2$^+$ cells in non-competitive (**k**) ($n = 10$ (control and *Tet2* knockout (*Tet2*KO)), $n = 8$ (*Srsf2$^{P95H}$*), and $n = 5$ (*Tet2*KO + *Srsf2$^{P95H}$*) at 0 weeks) or competitive (**l**) ($n = 9$ (control), $n = 10$ (*Tet2*KO), $n = 8$ (*Srsf2$^{P95H}$*), and $n = 10$ (*Tet2*KO + *Srsf2$^{P95H}$*) at 0 weeks) transplantation (pIpC was injected at 4 weeks post-transplant; percentages of CD45.2$^+$ cells at pre-transplant are also shown as data at 0 weeks in **l**; mean ± s.d.; two-way ANOVA with Tukey's multiple comparison test). **m**, **n**, Absolute number of BM HSPCs from two tibias, two femurs, and two pelvic bones were measured in the primary competitive transplant of *Tet2* and *Srsf2* mutant cells ($n = 10$ per genotype; mean + s.d.; two-way ANOVA with Tukey's multiple comparison test). **o**, Schematic of TET2 catalytic domain (CD: catalytic domain; EV: empty vector) retroviral BM transplantation model. **p**, Western blot analysis confirming the expression of Myc-tagged TET2 CD in Ba/F3 cells transduced with or without TET2 CD (representative images from two biologically independent experiments with similar results). **q**, Chimerism of mCherry–TET2 CD$^+$ and GFP–EV$^+$ cells in peripheral blood of recipient mice over time ($n = 10$; mean percentage ± s.d.; two-way ANOVA with Sidak's multiple comparison test). **r**, qPCR of *Tet3* in the first colony cells from **s** ($n = 3$; mean ± s.d.; a two-sided Student's *t*-test). **s**, Colony numbers from serial replating assays of BM cells from *Mx1-cre* control, *Mx1-cre Srsf2$^{P95H/+}$*, and *Mx1-cre Tet2$^{fl/fl}$Srsf2$^{P95H/+}$* mice transduced with shRNAs targeting *Tet3* (shTet3) ($n = 3$; mean + s.d.; two-way ANOVA with Tukey's multiple comparison test). **t**, Schematic of shTet3 retroviral BM transplantation model. **u**, **v**, Chimerism of mCherry$^+$ cells in CD45.2$^+$ donor cells in peripheral blood of recipient mice over time (**u**; left, *Mx1-cre Srsf2$^{P95H/+}$*; right, *Mx1-cre Tet2$^{fl/fl}$Srsf2$^{P95H/+}$*; $n = 5$ per group) and at 20 weeks post-transplant (**v**) (mean percentage ± s.d.; two-way ANOVA with Sidak's multiple comparison test). **w**, Colony numbers from serial replating assays of either *Mx1-cre Srsf2$^{+/+}$* or *Srsf2$^{P95H/+}$* BM cells transduced with an shRNA against *Fto* or *Alkbh5*. BM cells were collected at one month post-pIpC ($n = 3$; mean value ± s.d.; two-way ANOVA with Tukey's multiple comparison test). **x**, qPCR of *Fto* or *Alkbh5* in Ba/F3 cells transduced with shRNAs targeting mouse *Fto* or *Alkbh5* ($n = 3$; mean value ± s.d.; one-way ANOVA with Tukey's multiple comparison test). *$P < 0.05$, **$P < 0.01$, ***$P < 0.001$.
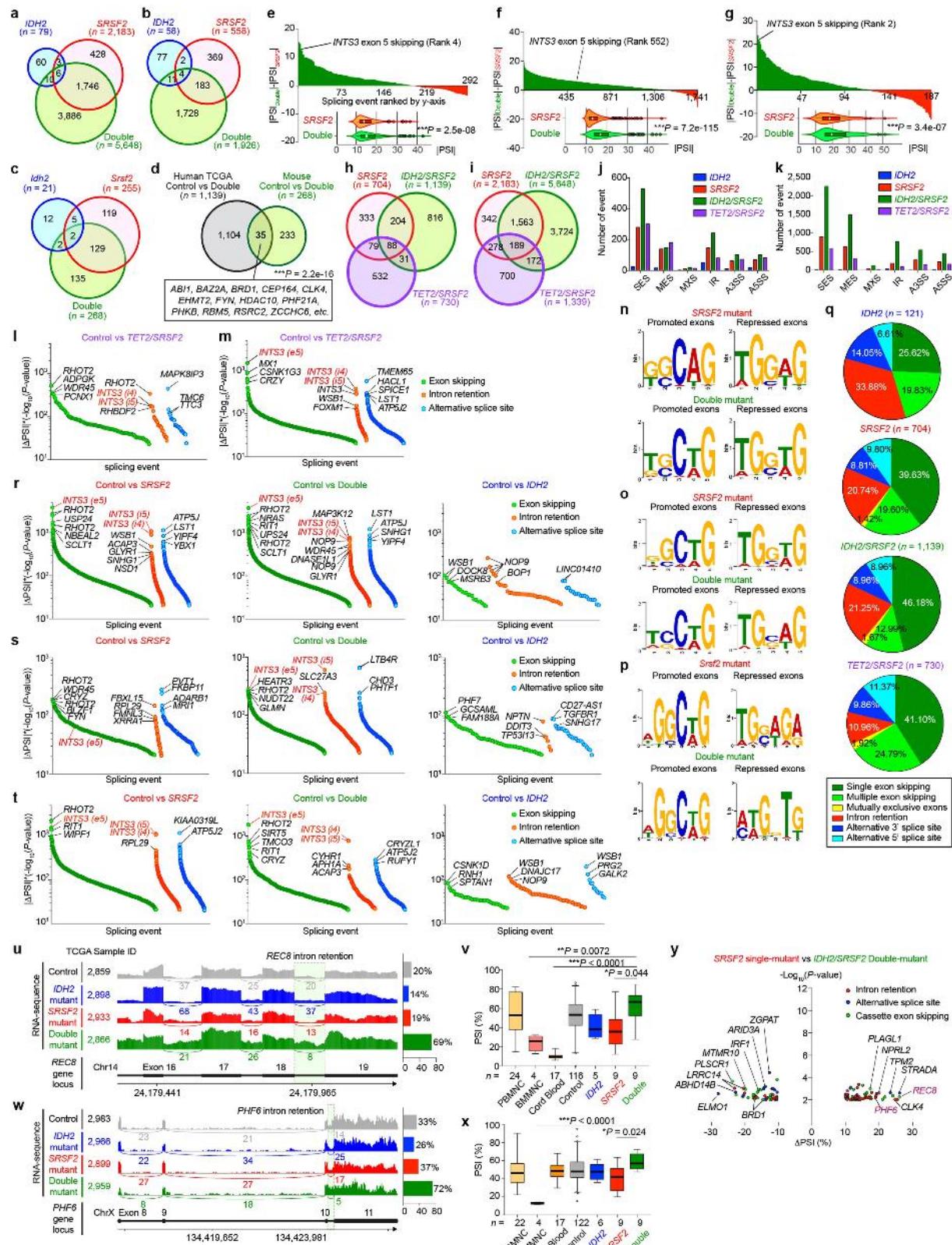
**Extended Data Fig. 5** | See next page for caption.

**Extended Data Fig. 5 | *IDH2* mutations augment the RNA splicing defects of *SRSF2* mutant leukaemia. a–c**, Venn diagram showing numbers of differentially spliced events from the Beat AML cohort (**a**), unpublished collaborative cohort 2 (**b**) and mouse Lin⁻KIT⁺ bone marrow cells at 12 weeks post-pIpC (**c**) based on *IDH2* and *SRSF2* mutant genotypes. **d**, Venn diagram showing the numbers of overlapping alternatively spliced events between *IDH2* and *SRSF2* double-mutant AMLs and mouse models (***$P = 2.2 \times 10^{-16}$; binominal test). **e–g**, $\Delta|\text{PSI}|$ ($\Delta|\text{PSI}| = |\text{PSI}|_{\text{Double}} - |\text{PSI}|_{SRSF2}$) values for each overlapping mis-spliced event in *SRSF2* single-mutant and *IDH2* and *SRSF2* double-mutant AML from the TCGA (**e**), Beat AML cohort (**f**) and unpublished collaborative cohort 2 (**g**) are plotted along the *y* axis. Spliced events shown in green and red represent events that are more robust in *IDH2* and *SRSF2* double-mutant and *SRSF2* single-mutant AML, respectively, in terms of |PSI| values. The mean |PSI| value of each event was visualized as violin plots on the bottom ($n = 292$, $n = 1,741$, and $n = 187$, respectively; PSI values were calculated using PSI-Sigma; the line represents mean, box edges show 25th and 75th percentiles and whiskers represent 2.5th and 97.5th percentiles; samples below 2.5th percentile and above 97.5th percentile are shown as dots; paired two-tailed Student *t*-test). **h, i**, Venn diagram of numbers of differentially spliced events from the TCGA (**h**) and Beat AML (**i**) datasets based on *IDH2*, *TET2* and *SRSF2* genotypes. **j, k**, Absolute numbers of each class of alternative splicing event from TCGA (**j**) and Beat AML (**k**) datasets are shown. SES, single-exon skipping; MES, multiple-exon skipping; MXS, mutually-exclusive splicing; A5SS, alternative 5′ splice site; A3SS, alternative 3′ splice site. **l, m**, Differentially spliced events ($|\Delta\text{PSI}| > 10\%$ and $P < 0.01$ were used as thresholds) in indicated genotype from the TCGA (**l**) ($n = 730$ differentially spliced events) and Beat AML (**m**) ($n = 1,339$ differentially spliced events) cohorts

are ranked by *y* axis and class of event (PSI and *P* values adjusted for multiple comparisons were calculated using PSI-Sigma). **n–p**, Sequence logos of nucleotide motifs of exons preferentially promoted or repressed in splicing in *SRSF2* single-mutant (top) or *IDH2* and *SRSF2* double-mutant (bottom) AML from the TCGA cohort (**n**), Beat AML cohort (**o**) and mouse models (**p**). **q**, Percentage of each class of alternative splicing event in indicated genotype from TCGA cohort is shown in a pie-chart. **r–t**, Differentially spliced events ($|\Delta\text{PSI}| > 10\%$ and $P < 0.01$ were used as thresholds) in indicated genotype from the Beat AML (**r**) ($n = 2,183$, 5,648, and 79 differentially spliced events, respectively), unpublished collaborative cohort 2 (**s**) ($n = 558$, 1,926, and 94 differentially spliced events, respectively) and Leucegene cohort (**t**) ($n = 2,571$, 787, and 122 differentially spliced events, respectively) are ranked by *y*-axis and class of event (PSI and *P* values adjusted for multiple comparisons were calculated using PSI-Sigma). **u, w**, Representative Sashimi plots of RNA-seq data showing the intron retention events in *REC8* (**u**) and *PHF6* (**q**) from the TCGA dataset. **v, x**, PSI values for intron retention events in *REC8* (**v**) and *PHF6* (**x**) in normal PBMNCs (GSE58335[41]), BMMNCs (GSE61410[41]), cord blood CD34⁺ cells (GSE48846[42]), and AML samples with indicated genotypes (the line represents the median, box edges show 25th and 75th percentiles and whiskers represent 2.5th and 97.5th percentiles; samples below 2.5th percentile and above 97.5th percentile are shown as dots; PSI and *P* values adjusted for multiple comparisons were calculated using PSI-Sigma; one-way ANOVA with Tukey's multiple comparison test; *$P < 0.05$; **$P < 0.01$; ***$P < 0.001$). **y**, Volcano plots of aberrant splicing events in TCGA AML data comparing *SRSF2* single-mutant and *IDH2*/*SRSF2* double-mutant AML ($n = 122$ differentially spliced events; PSI and *P* values adjusted for multiple comparisons were calculated using PSI-Sigma; $|\Delta\text{PSI}| > 10\%$ and $P < 0.01$ were used as thresholds).
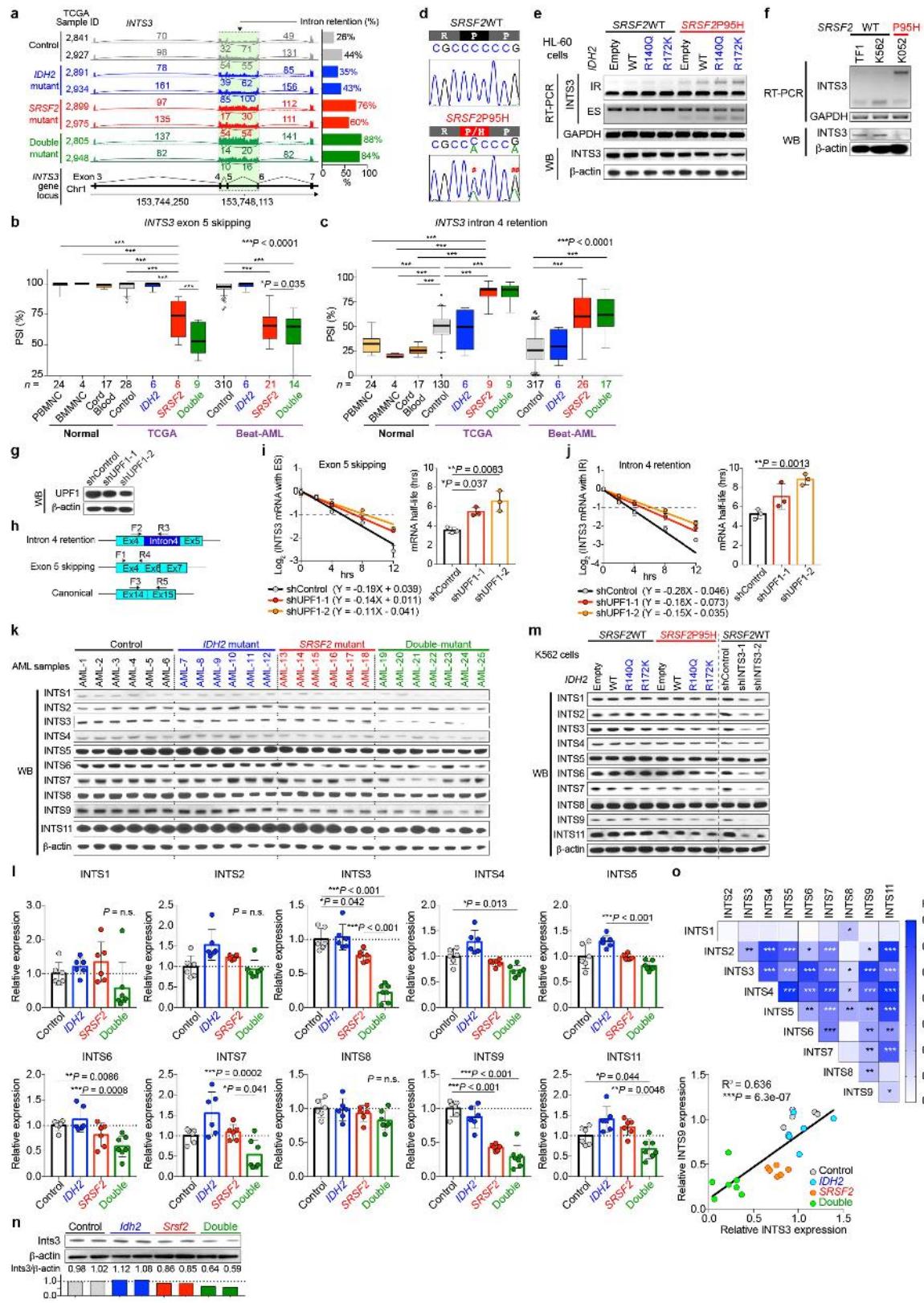
**Extended Data Fig. 6 |** See next page for caption.

**Extended Data Fig. 6 | Aberrant *INTS3* transcripts undergo nonsense-mediated decay, and effect of INTS3 loss extends to other members of the integrator complex. a**, Representative Sashimi plots of RNA-seq data from the TCGA showing intron retention in *INTS3*. **b, c**, PSI values for *INTS3* exon 5 skipping (**b**) and intron 4 retention (**c**) in normal PBMNC (GSE58335[41]), BMMNC (GSE61410[41]), cord blood CD34+ cells (GSE48846[42]) and AML samples with indicated genotypes (the number of RNA-seq samples analysed is indicated; PSI and *P* values adjusted for multiple comparisons were calculated using PSI-Sigma; the line represents mean, box edges show 25th and 75th percentiles and whiskers represent 2.5th and 97.5th percentiles; samples below 2.5th percentile and above 97.5th percentile are shown as dots; one-way ANOVA with Tukey's multiple comparison test). **d**, Sanger sequencing of cDNA showing wild-type or mutant *SRSF2* expression in isogenic K562 knock-in cells. [#]a nonsynonymous mutation that alters P95; [##]a synonymous mutation that does not change the amino acid. **e**, RT–PCR and western blot analysis of INTS3 in isogeneic HL-60 cells with various combinations of *IDH2* and *SRSF2* mutations. IR: intron retention; ES: exon skipping. Representative results from three biologically independent experiments with similar results. **f**, RT–PCR and western blot of INTS3 in non-isogenic myeloid leukaemia cell lines. *SRSF2* genotypes are shown together (representative results from three independent experiments with similar results). **g**, Western blot analysis of K562 *SRSF2^P95H^* knock-in cells transduced with shRNAs against *UPF1* (representative results from three biologically independent experiments with similar results). **h**, Primers used to specifically measure *INTS3* isoform with intron 4 retention and exon 5 skipping, and those for the normal *INTS3* isoform. **i, j**, Half-life of *INTS3* transcripts with exon 5 skipping (**i**) and intron 4 retention (**j**) were measured by qPCR ($n = 3$; mean ± s.d.; a two-sided Student's *t*-test). **k, l**, Western blot analysis of protein lysates of samples from patients with AML with the indicated *IDH2* and *SRSF2* genotypes (**k**). Expression level of each integrator subunit was quantified using ImageJ and relative expression levels are shown in **l**, in which the mean expression levels of control samples were set as 1 ($n = 6$ for control, *IDH2* single-mutant, and *SRSF2* single-mutant AML, and $n = 7$ for *IDH2* and *SRSF2* double-mutant AML; detailed information of the primary patient samples used for this analysis is provided in Supplementary Table 23; mean ± s.d.; one-way ANOVA with Tukey's multiple comparison test). **m**, Western blot analysis of protein lysates from isogenic K562 cells with indicated *IDH2* and *SRSF2* genotypes (left) or with INTS3 knockdown (right) (representative results from three biologically independent experiments are shown). **n**, Western blot analysis of murine Lin−KIT+ BM cells at 12 weeks post-pIpC based on *Idh2* and *Srsf2* mutant genotypes. Expression level of INTS3 was quantified using ImageJ and relative expression levels are shown below; $n = 2$ mice per genotype were analysed. **o**, Correlation among indicated Integrator subunits and *P* value were calculated in Excel and $R^2$ values are visualized as a heat map generated by Prism 7 (top). Correlation between INTS3 and INTS9 protein expression is shown (bottom) ($n = 25$ from **k**; the Pearson correlation coefficient ($R^2$) and *P* values (two-tailed) were calculated in Excel). *$P < 0.05$, **$P < 0.01$, ***$P < 0.001$.
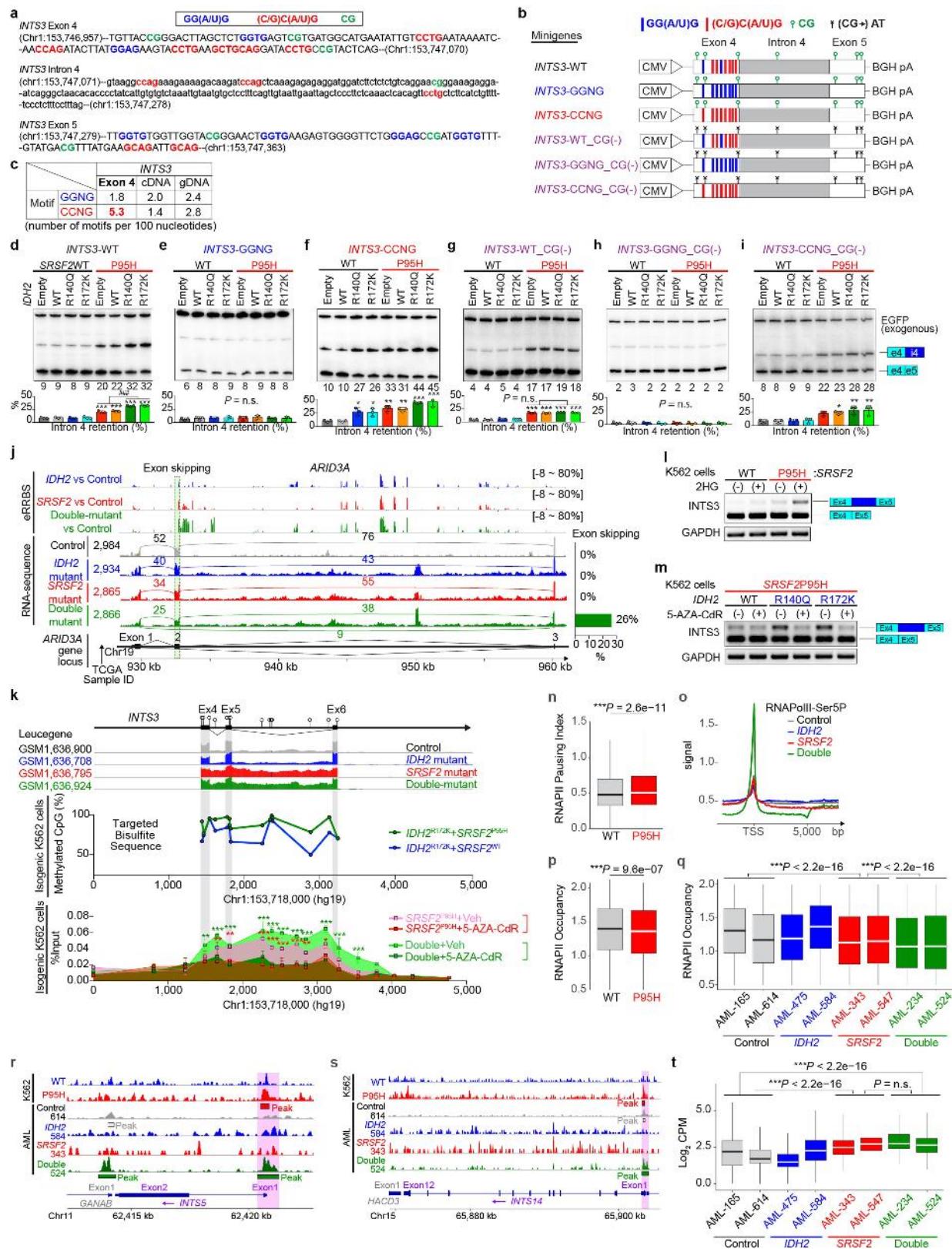
**Extended Data Fig. 7 |** See next page for caption.

**Extended Data Fig. 7 | DNA hypermethylation at *INTS3* enhances *INTS3* mis-splicing, which is associated with RNAPII stalling.**
**a**, Sequence of human *INTS3* exon 4, intron 4 and exon 5, and schematic of *INTS3* minigene constructs. GG(A/U)G motifs, (C/G)C(A/U)G motifs, and CG dinucleotides are highlighted in blue, red, and green, respectively. **b**, Schematic of *INTS3* minigene constructs. **c**, Table revealing the number of GGNG or CCNG motifs in exon 4, entire cDNA of *INTS3* or entire genomic DNA (gDNA) of *INTS3* per 100 nucleotides. **d**–**i**, Radioactive RT–PCR results of *INTS3* minigene assays using indicated versions of the minigene in isogenic K562 cells. Percentage of intron 4 retention were normalized against exogenous eGFP ($n = 3$; mean percentage ± s.d.; one-way ANOVA with Tukey's multiple comparison test). **j**, Mean percentage of methylated CpGs at *ARID3A* in samples from patients with AML with indicated genotypes determined by eRRBS ($n = 3$ patients per genotype), followed by IGV plots of RNA-seq data of *ARID3A* from the TCGA. **k**, Results of eRRBS ($n = 1$ per genotype) and RNAPII-Ser2P ChIP-walking experiments are represented as shown in Fig. 3f ($n = 3$; mean percentage ± s.d.; two-way ANOVA with Tukey's multiple comparison test). **l**, **m**, RT–PCR results detecting *INTS3* intron retention in isogenic K562 cells containing various combinations of *IDH2* and *SRSF2* mutations that were treated with cell-permeable 2HG at 0.5 μM (**l**) or 5-AZA-CdR at 5 μM (**m**) for 8 days (representative results from three biologically independent experiments with similar results). **n**, RNAII pausing index in isogenic *SRSF2*$^{WT}$ or *SRSF2*$^{P95H}$ mutant K562 cells was calculated as previously described[20] as a ratio of normalized ChIP–seq reads of RNAPII-Ser5P on TSSs (±250 bp) over that of the corresponding bodies (+500 to +1,000 from TSSs) (the line represents the median, box edges show 25th and 75th percentiles and whiskers represent 2.5th and 97.5th percentiles; each box plot was made by analysing ChIP–seq data from one cell line; two-sided Student's *t*-test). **o**, Metagene plots showing genome-wide RNAPII-Ser5P occupancy in primary samples from patients with AML with indicated genotypes (patient samples used for this analysis are described in Supplementary Table 23). **p**, **q**, RNAPII occupancy representing ChIP–seq reads of RNAPII-Ser2P over gene bodies was calculated for isogenic K562 cells (**p**) and AML samples (**q**) (the line represents the median, box edges show 25th and 75th percentiles and whiskers represent 2.5th and 97.5th percentiles; each box plot was made by analysing ChIP–seq data from one cell line (**p**) or one primary AML sample (**q**); two-sided Student's *t*-test (**p**) and one-way ANOVA with Tukey's multiple comparison test (**q**)). **r**, **s**, Genome browser view of ChIP–seq signal for RNAPII Ser5P at *INTS5* (**r**) and *INTS14* (**s**) in isogenic K562 cells with or without *SRSF2* mutation ($n = 1$) and primary AML samples with indicated genotype (results generated from $n = 2$ primary AML samples are shown). **t**, RNAPII abundance over the differentially spliced regions between *IDH2* and *SRSF2* wild-type control and *SRSF2* single-mutant AML determined by RNAPII-Ser2P ChIP–seq (*y* axis, log$_2$(counts per million); the line represents the median, box edges show 25th and 75th percentiles and whiskers represent 2.5th and 97.5th percentiles; each box plot was made by analysing ChIP–seq data from one primary AML sample; one-way ANOVA with Tukey's multiple comparison test. *$P < 0.05$, **$P < 0.01$, ***$P < 0.001$.
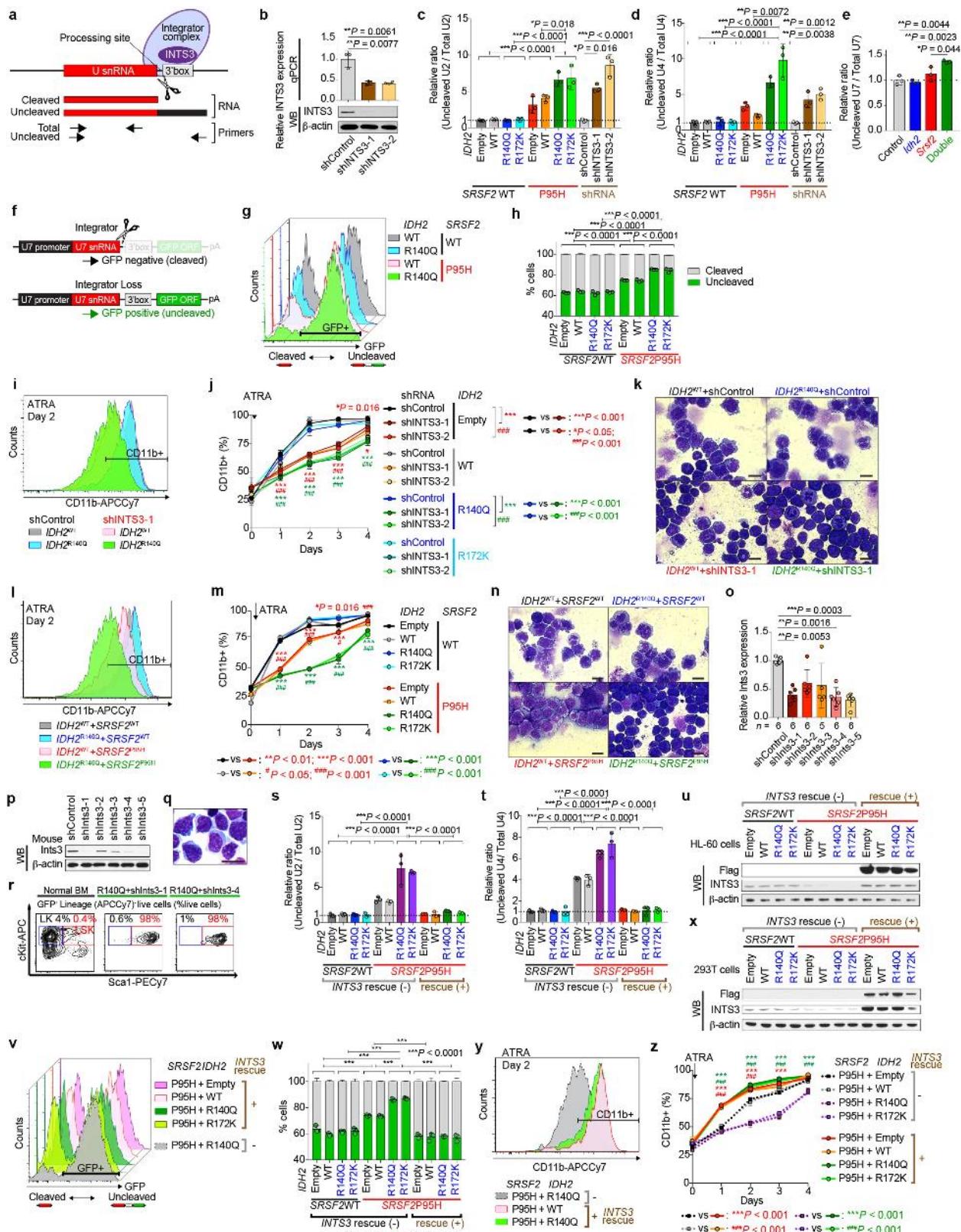
**Extended Data Fig. 8 |** See next page for caption.

**Extended Data Fig. 8 | Loss of INTS3 impairs uridine-rich small nuclear RNA processing and blocks myeloid differentiation.**
**a**, Schematic of snRNA processing site and qPCR primers for detecting cleaved or uncleaved snRNA. **b**, qPCR (top; $n = 3$; mean ± s.d.; a two-sided Student's $t$-test) and representative western blot of INTS3 in HL-60 cells transduced with shRNAs targeting human *INTS3* (bottom, representative results from three biologically independent experiments). **c**–**e**, **s**, **t**, qPCR results of U2 (**c**, **s**) and U4 (**d**, **t**) snRNAs in isogenic HL-60 cells and U7 snRNA in murine cells from Extended Data Fig. 6n (**e**). Ratio of uncleaved/total snRNAs expression was compared ($n = 3$, mean ratio ± s.d.; one-way ANOVA with Tukey's multiple comparison test; the largest $P$ values calculated among $2 \times 2$ comparisons of two components from different groups are shown. For example, $P$ values were calculated from the following four comparisons; bars 1 versus 3, 2 versus 3, 1 versus 4, 2 versus 4). **f**, Schematic of the U7 snRNA–GFP reporter. **g**, **v**, Flow cytometry analysis of 293T cells transduced with U7 snRNA-GFP reporter and *IDH2*, *SRSF2* and *INTS3* constructs as labelled on the right (representative results from three biologically independent experiments are shown). **h**, **w**, Quantification of per cent GFP⁻ and GFP⁺ 293T cells ($n = 3$ biologically independent experiments, mean percentage ± s.d.; one-way ANOVA with Tukey's multiple comparison test; $P$ values are shown as in **c**). **i**, **l**, **y**, Flow cytometry analysis of CD11b expression in isogenic HL-60 cells after ATRA treatment for two days (representative results from three biologically independent experiments are shown). **j**, **m**, **z**, Quantification of percentages of CD11b⁺ HL-60 cells over time ($n = 3$; mean percentage ± s.d.; two-way ANOVA with Tukey's multiple comparison test). **k**, **n**, Cytomorphology of isogenic HL-60 cells after ATRA treatment for two days (Giemsa staining; scale bar, 10 μm; original magnification, ×400; representative results from three biologically independent experiments are shown). **o**, **p**, qPCR of *Ints3* (**o**) (mean ± s.d.; Kruskal–Wallis tests with uncorrected Dunn's test) and western blot of INTS3 (**p**) in Ba/F3 cells transduced with shRNAs targeting mouse *Ints3*. **q**, **r**, Representative cytomorphology (**q**) and immunophenotype (**r**) of colony cells at the sixth colony. Normal BMMNCs were used as a control (the percentage listed represent the percent of cells within live cells; representative results from three biologically independent experiments are shown). **u**, **x**, Western blot of proteins extracted from HL-60 cells (**u**) assayed in **s-t** and **y-z** and 293T cells (**x**) assayed in **v** and **w** (representative results from three biologically independent experiments). *$P < 0.05$, **$P < 0.01$, ***$P < 0.001$, #$P < 0.05$, ##$P < 0.01$, ###$P < 0.001$.
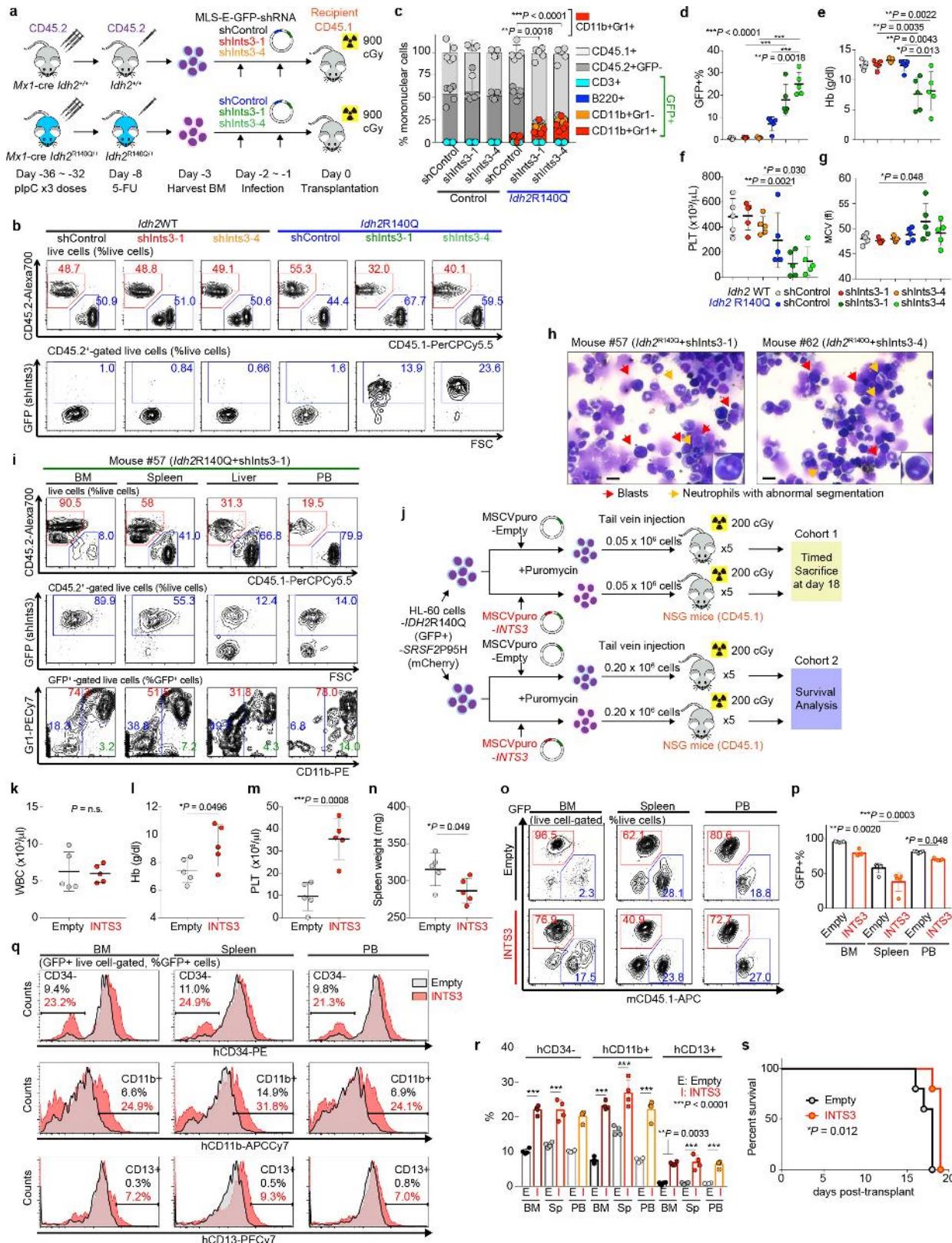
**Extended Data Fig. 9 |** See next page for caption.

**Extended Data Fig. 9 | Mutant Idh2 cooperates with Ints3 loss to generate a lethal myeloid neoplasm in vivo. a**, Schematic of shRNA targeting *Ints3* (shInts3) retroviral BM transplantation model. **b**, Flow cytometry data showing the chimerism of CD45.2$^+$ versus CD45.1$^+$ (top) or GFP$^+$ (bottom) cells in peripheral blood at four weeks post-transplant (the percentages listed represent the percent of cells within live cells; representative results from five recipient mice). **c**, Composition of PBMNCs at four weeks post-transplant ($n = 5$ per group; mean + s.d.; represented by lines above the box. statistical significance was detected in percentage of CD11b$^+$Gr1$^+$ cells; by two-way ANOVA with Tukey's multiple comparison test). **d**–**g**, Chimerism of GFP$^+$ cells in peripheral blood (**d**) and blood counts of recipients at four weeks post-transplant (Hb (**e**); PLT (**f**); MCV (**g**); $n = 5$ per group; mean ± s.d.; one-way ANOVA with Tukey's multiple comparison test). **h**, Giemsa staining of BMMNCs from moribund mice with indicated genotypes (red and yellow arrows represent blastic cells and dysplastic neutrophils, respectively; inset, representative neutrophils with abnormal segmentation; scale bar, 10 μm; original magnification, ×400; representative results from five mice per genotype). **i**, Flow cytometry data of BM, spleen, liver, and peripheral blood from *Idh2$^{R140Q}$* mice treated with shInts3 (representative results from five mice). **j**, Schematic of HL-60 xenograft model in which recipient mice from cohort 1 were euthanized at day 18 post-transplant and mice from cohort 2 were observed for survival analysis until end stage. **k**–**n**, Blood counts (WBC (**k**); Hb (**l**); PLT (**m**)) and spleen weight (**n**) of mice from cohort 1 at day 18 post-transplant (mean ± s.d.; $n = 5$ per group; a two-sided Student's *t*-test). **o**, **p**, Representative flow cytometry data of BM, spleen, and peripheral blood from the recipient mice from cohort 1 (**o**) (the percentage represents the percent of cells within live cells) and the mean percentage of GFP$^+$ cells (**p**) ($n = 5$ per group; mean + s.d.; two-way ANOVA with Sidak's multiple comparison test). **q**, **r**, Representative flow cytometry data of BM, spleen and peripheral blood from cohort 1 (**q**) (the percentage represents the percent of cells within GFP$^+$ live cells) and the mean percentage of hCD34$^-$, hCD11b$^+$ and hCD13$^+$ cells (**r**) ($n = 4$ per group; mean + s.d.; two-way ANOVA with Sidak's multiple comparison test). **s**, Kaplan–Meier survival analysis of recipient mice from cohort 2 ($n = 5$ per group; log-rank (Mantel-Cox) test (two-sided)). *$P < 0.05$, **$P < 0.01$, ***$P < 0.001$.
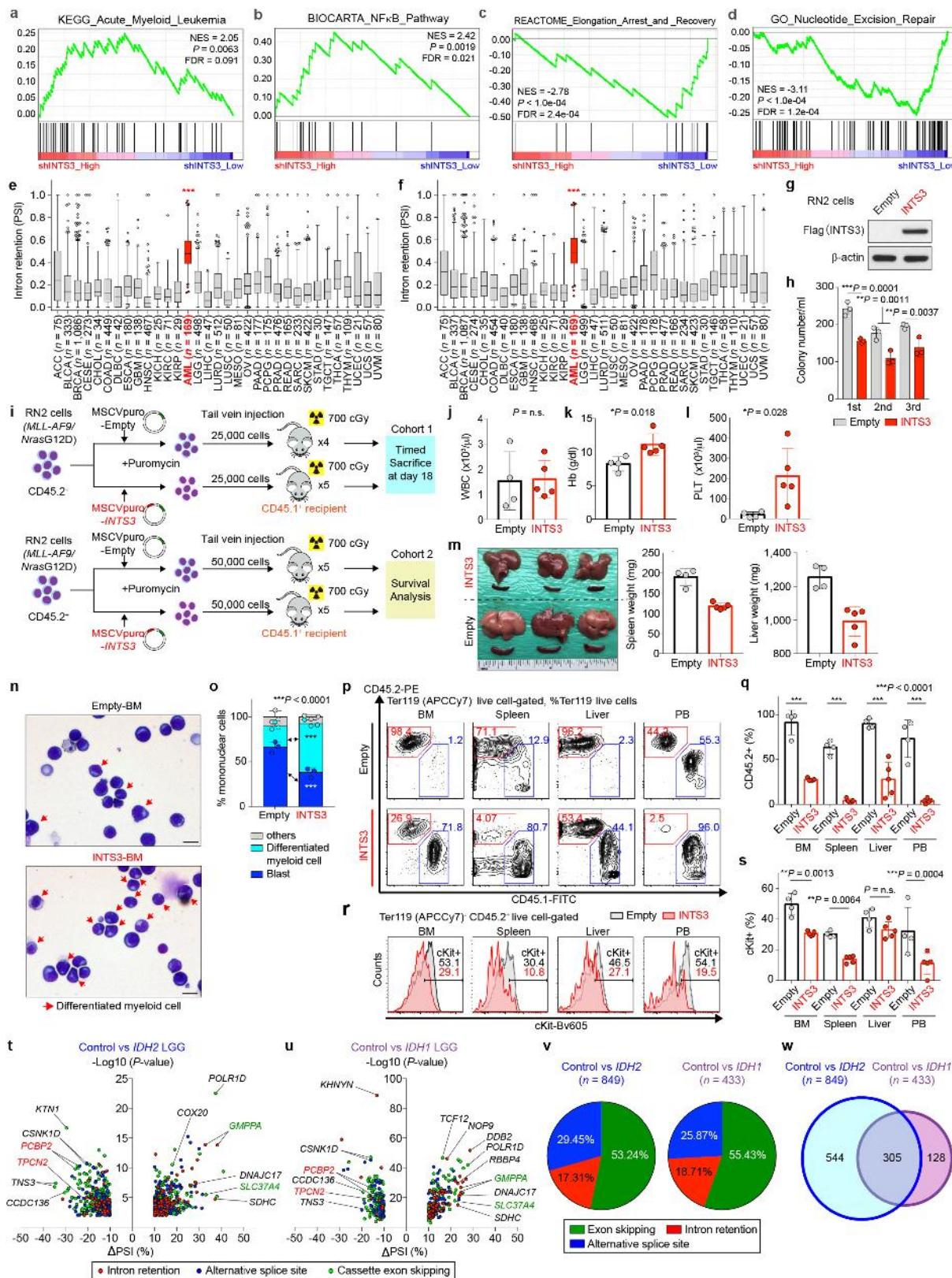
**Extended Data Fig. 10 |** See next page for caption.

**Extended Data Fig. 10 | Gene expression and biological consequences of INTS3 loss, and effect of *IDH1* and *IDH2* mutations on splicing in low-grade glioma. a–d**, GSEA based on RNA-seq data generated from isogenic *IDH2^{R140Q}* mutant HL-60 cells with or without INTS3 depletion. Representative results from gene sets associated with leukaemogenesis and myeloid differentiation (**a**), oncogenic signalling pathways (**b**), RNAPII elongation-linked transcription (**c**) and DNA damage response (**d**) with statistical significance ($P < 0.01$) are shown (*y* axis; enrichment score; NES: normalized enrichment score; FDR: false discovery rate; RNA-seq data generated from isogenic HL-60 cells in duplicate were analysed using GSEA[34]). **e, f**, PSI values for *INTS3* intron 4 (**e**) and 5 (**f**) retention events across 33 cancer cell types (the same datasets were analysed in Fig. 4f). ACC,: adrenocortical carcinoma; BLCA, bladder urothelial carcinoma; BRCA, breast invasive carcinoma; CESC, cervical squamous cell carcinoma and endocervical adenocarcinoma; CHOL, cholangiocarcinoma; DLBC, diffuse large B-cell lymphoma; ESCA, oesophageal carcinoma; GBM, glioblastoma multiforme; HNSC, head and neck squamous cell carcinoma; KICH, kidney chromophobe; KIRC, kidney renal clear cell carcinoma; KIRP, kidney renal papillary cell carcinoma; LGG, low-grade glioma; LIHC, liver hepatocellular carcinoma; LUSC, lung squamous cell carcinoma; MESO, mesothelioma; OV, ovarian serous cystadenocarcinoma; PRAD, prostate adenocarcinoma; READ, rectum adenocarcinoma; SARC, sarcoma; SKCM, skin cutaneous melanoma; STAD, stomach adenocarcinoma; TGCT, testicular germ cell tumours; THCA, thyroid carcinoma; THYM, thymoma; UCEC, uterine corpus endometrial carcinoma; UCS, uterine carcinosarcoma; UVM, uveal melanoma. The line represents the median, box edges show 25th and 75th percentiles and whiskers represent 2.5th and 97.5th percentiles; samples below 2.5th percentile and above 97.5th percentile are shown as dots; one-way ANOVA with Dunnett's multiple comparison test; ***$P < 0.001$ represents the *P* values from all the comparisons between AML and any of other 32 non-AML cancer type. **g**, Western blot analysis confirming overexpression of 3× Flag-tagged INTS3 in RN2 (*MLL-AF9 Nras^{G12D}*) leukaemia cells (representative results from three biologically independent experiments). **h**, Colony numbers from serial replating assays of RN2 cells with or without INTS3 overexpression ($n = 3$; mean + s.d. represented by lines above the box; two-way ANOVA with Sidak's multiple comparison test). **i**, Schematic of INTS3 retroviral BM transplantation models in which recipient mice from cohort 1 were euthanized at day 18 post-transplant and mice from cohort 2 were observed for survival analysis until end-stage. **j–l**, Blood counts (WBC (**j**); Hb (**k**); PLT (**l**)) of mice from cohort 1 at day 18 post-transplant (mean ± s.d.; $n = 4$ ('empty' group); $n = 5$ ('INTS3' group) recipient mice; a two-sided Student's *t*-test). **m**, Representative photograph of spleens and livers from cohort 1 with an inch scale (left), and spleen (middle) and liver weight (right) ($n = 4$ (empty); $n = 5$ (INTS3); mean ± s.d.; two-sided Student's *t*-test). **n, o**, Representative Giemsa staining (**n**) (red arrows represent differentiated cells; scale bar, 10 μm; original magnification, ×400) and percentages of blasts, differentiated myeloid cells, and other cells in BMMNCs (**o**) from moribund mice from cohort 2 ($n = 3$ per genotype; 100 cells per mouse were classified; mean percentage + s.d.; two-way ANOVA with Sidak's multiple comparison test). **p, q**, Representative flow cytometry analysis of BM, spleen, liver, and peripheral blood (**p**) and percentages of CD45.2^+ cells in Ter119^− live cells (**q**) in recipient from cohort 1 ($n = 4$ (empty); $n = 5$ (INTS3); mean ± s.d.; two-way ANOVA with Tukey's multiple comparison test). **r, s**, Representative flow cytometry analysis showing KIT expression in RN2 cells with or without INTS3 overexpression (**r**) and quantification of KIT^+ cells (**s**) from cohort 1 ($n = 4$ (Empty); $n = 5$ (INTS3); mean ± s.d.; one-way ANOVA with Tukey's multiple comparison test). **t, u**, Volcano plots of aberrant splicing events in the LGG TCGA dataset based on *IDH2* (**t**) or *IDH1* (**u**) mutant genotypes. |ΔPSI| > 10% and $P < 0.01$ were used as thresholds ($n = 849$ and $n = 433$ differentially spliced events, respectively; RNA-seq data were analysed using PSI-Sigma). **v**, Percentage of each class of alternative splicing event in *IDH2* (left) and *IDH1* (right) mutant LGG is shown in pie-chart. **w**, Venn diagram of numbers of alternatively spliced events from the LGG TCGA dataset based on *IDH1* and *IDH2* mutant genotypes. 'Control' represents LGG with wild-type *IDH1* and *IDH2*. *$P < 0.05$, **$P < 0.01$, ***$P < 0.001$.

# naturers research

Corresponding author(s):    Omar Abdel-Wahab

Last updated by author(s):   Jul 17, 2019

# Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

## Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

| n/a | Confirmed | |
|---|---|---|
| ☐ | ☒ | The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement |
| ☐ | ☒ | A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly |
| ☐ | ☒ | The statistical test(s) used AND whether they are one- or two-sided<br>*Only common tests should be described solely by name; describe more complex techniques in the Methods section.* |
| ☐ | ☒ | A description of all covariates tested |
| ☐ | ☒ | A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons |
| ☐ | ☒ | A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals) |
| ☐ | ☒ | For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted<br>*Give P values as exact values whenever suitable.* |
| ☒ | ☐ | For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings |
| ☒ | ☐ | For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes |
| ☒ | ☐ | Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated |

*Our web collection on statistics for biologists contains articles on many of the points above.*

## Software and code

Policy information about availability of computer code

| Data collection | No code was used for data collection. |
|---|---|
| Data analysis | The inclusion ratios of alternative exons or introns were estimated by using PSI-Sigma. GraphPad Prism 7 software was used to analyze the data and to make figures. RNA-seq reads were aligned by using 2-pass STAR 2.5.2a. Samtools (1.3.1) were used to generate variant call format (VCF) files for 6 target genes: IDH1, IDH2, SF3B1, SRSF2, U2AF1, and ZRSR2 with mpileup parameters (-Bvu). The VCF files were further processed by our in-house scripts to filter out mutations whose VAF was lower than 15%. The filtered VCF files were used for variant effect predictor (version 89.4) to annotate the consequences of the mutations. Motif analysis was done by using MEME SUITE. The heatmaps and sample clustering were done by using MORPHEUS (software.broadinstitute.org/morpheus/).<br>VCF files from the TruSightTM Myeloid 54 gene panel from Illumina were analyzed using Illumina's Variant Studio software while those from a 40 gene panel (Oncomine Myeloid Research Assay; ThermoFisher), processing eight samples per Ion 530 chip on the IonTorrent platform were analyzed using the Ion Reporter software.<br>ChIP-seq reads were mapped to the genome by calling Bowtie v1.0.048 with the arguments '-v 2 -k 1 -m 1 --best --strata'. Peaks were called using MACS2 v2.1.1.2016030952 against input control libraries with P < 1e-5 and subsequently filtered to remove peaks contained within ENCODE blacklisted regions and the mitochondrial genome. Subsequent data analysis was performed with Bioconductor in the R programming environment. Consensus peaks between samples were called using the soGGI package v1.14.0. Peaks were annotated using the ChIPseeker package v1.18.0. |

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

# Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

The data that support the findings of this study are available from the corresponding author upon reasonable request. The RNA sequencing data have been deposited in NCBI Sequencing Read Archive (SRA) under accession number SRP133673.

# Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences    ☐ Behavioural & social sciences    ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](#)

# Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

| | |
|---|---|
| Sample size | For the in vivo experiments, the number of mice in each experiment was chosen to provide 90% statistical power with a 5% error level. For the RNA-seq experiments, maximal sample sizes available were used to obtain statistical power as much as possible to detect significant splicing alterations. |
| Data exclusions | No data were excluded from the analyses. |
| Replication | The experiments were repeated at a minimum of 3 times for all the in vitro experiments. All attempts at replication were successful. |
| Randomization | Animals were assigned to experimental group based on genotype and there was no drug treatment groups, therefore randomization was not utilized. |
| Blinding | For survival and blood count analyses of mice, actual measurements were carried out by a member of the lab who did not have knowledge of which alleles were expected to alter survival or blood count parameters. All other experiments were not blinded and it was not necessary to be as they were less subjective. |

# Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

## Materials & experimental systems

| n/a | Involved in the study |
|---|---|
| ☐ | ☒ Antibodies |
| ☐ | ☒ Eukaryotic cell lines |
| ☒ | ☐ Palaeontology |
| ☐ | ☒ Animals and other organisms |
| ☐ | ☒ Human research participants |
| ☒ | ☐ Clinical data |

## Methods

| n/a | Involved in the study |
|---|---|
| ☐ | ☒ ChIP-seq |
| ☐ | ☒ Flow cytometry |
| ☒ | ☐ MRI-based neuroimaging |

# Antibodies

| | |
|---|---|
| Antibodies used | Flow cytometry antibodies: B220-APCCy7 (clone: RA3-6B2; purchased from BioLegend; catalog #: 103224; dilution: 1:200); B220-Bv711 (RA3-6B2; BioLegend; 103255; 1:200); CD3-PerCPCy5.5 (17A2; BioLegend; 100208; 1:200); CD3-APC (17A2; BioLegend; 100236; 1:200); CD3-APCCy7 (17A2; BioLegend; 100222; 1:200); Gr1-PECy7 (RB6-8C5; eBioscience; 25-5931-82; 1:500); CD11b-PE (M1/70; eBioscience; 12-0112-85; 1:500); CD11b-APCCy7 (M1/70; BioLegend; 101226; 1:200); CD11c-APCCy7 (N418; BioLegend; 117323; 1:200); NK1.1-APCCy7 (PK136; BioLegend; 108724; 1:200); Ter119-APCCy7 (BioLegend; 116223; 1:200); cKit-APC (2B8; BioLegend; 105812; 1:200); cKit-PerCPCy5.5  (2B8; BioLegend; 105824; 1:100); cKit-Bv605 (ACK2; BioLegend; 135120; 1:200); Sca1-PECy7 (D7; BioLegend; 108102; 1:200); CD16/CD32 (FcγRII/III)-Alexa700 (93; eBioscience; 56-0161-82; 1:200); CD34-FITC (RAM34; BD Biosciences; 553731; 1:200); CD45.1-FITC (A20; BioLegend; 110706; 1:200); CD45.1-PerCPCy5.5 (A20; BioLegend; 110728; 1:200); CD45.1-PE (A20; BioLegend; 110708; 1:200); CD45.1-APC (A20; BioLegend; 110714; 1:200); CD45.2- |

PE (104; eBioscience; 12-0454-82; 1:200); CD45.2-Alexa700 (104; BioLegend; 109822; 1:200); CD45.2-Bv605 (104; BioLegend; 109841; 1:200); CD48-Bv711 (HM48-1; BioLegend; 103439; 1:200); CD150 (9D1; eBioscience; 12-1501-82; 1:200); CD34-PerCP (8G12; BD Biosciences; 345803; 1:200); CD117-PECy7 (104D2; eBioscience; 25-1178-42; 1:200); CD33-APC (P67.6; BioLegend; 366606; 1:200); HLA-DR-FITC (L243; BioLegend; 307604; 1:200); CD13-PE (L138; BD Biosciences; 347406; 1:200); CD45-APC-H7 (2D1; BD Biosciences; 560178; 1:200).

Western blotting, DNA dot blot assays, and ChIP: INTS1 (purchased from Bethyl laboratories; catalog #: A300-361A; dilution: 1:1,000), INTS2 (Abcam; ab74982; 1:1,000), INTS3 (Bethyl laboratories; A300-427A; 1:1,000, Abcam; ab70451; 1:1,000), INTS4 (Bethyl laboratories; A301-296A; 1:1,000), INTS5 (Abcam; ab74405; 1:1,000), INTS6 (Abcam; ab57069; 1:1,000), INTS7 (Bethyl laboratories; A300-271A; 1:1,000), INTS8 (Bethyl laboratories; A300-269A; 1:1,000), INTS9 (Bethyl laboratories; A300-412A; 1:1,000), INTS11 (Abcam; ab84719; 1:1,000), Flag-M2 (Sigma-Aldrich; F-1084; 1:1,000), Myc-tag (Cell Signaling; 2276S; 1:1,000), $\beta$-actin (Sigma-Aldrich; A-5441; 1:2,000), 5-Hydroxymehylcytosine (5hmC) (Active motif; 39769), RNA polymerase II CTD repeat YSPTSPS (phospho S2) (Abcam; ab5095), RNA polymerase II CTD repeat YSPTSPS (phospho S5) (Abcam; ab5408), and UPF1 (Abcam; ab109363; 1:1,000).

| Validation | All antibodies were validated by the supplier for human samples, and were checked in the lab by Western blotting on cell lysate and by comparing to the manufacturer's or in-house results. |

## Eukaryotic cell lines

Policy information about cell lines

| Cell line source(s) | K562, HL-60, TF1, and HEK293T cells were obtained from the American Type Culture Collection (ATCC) . K052 cells were obtained from JCRB Cell Bank. Ba/F3 cells were obtained from DSMZ. 293 GPII cells were purchased from Clontech. The isogenic K562 cell lines with or without SRSF2 P95H were generated at Horizon Discovery. MLL-AF9/NrasG12D murine leukemia (RN2) cells were obtained from Dr. Iannis Aifantis (NYU School of Medicine). |
| --- | --- |
| Authentication | An aliquot of each cell lines were authenticated using ATCC/JCRB/DSMZ DNA fingerprinting. |
| Mycoplasma contamination | All cell lines are frequently tested for mycoplasma contamination. Cell lines used in this study were verified to be mycoplasma negative before undertaking any experiments with them. |
| Commonly misidentified lines (See ICLAC register) | No commonly misidentified cell lines were used. |

## Animals and other organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research

| Laboratory animals | 6-8 week female CD45.1 C57BL/6 mice were purchased from The Jackson Laboratory (Stock No: 002014) . Male and female CD45.2 Srsf2P95H/+ conditional knock-in mice, Idh2R140Q/+ conditional knock-in mice, and Tet2 conditional knockout mice (all on C57BL/6 background) were also analyzed and used as bone marrow donors. |
| --- | --- |
| Wild animals | The study did not involve wild animals. |
| Field-collected samples | The study did not involve samples collected from the field. |
| Ethics oversight | All animal procedures were completed in accordance with the Guidelines for the Care and Use of Laboratory Animals and were approved by the Institutional Animal Care and Use Committees at MSKCC. |

Note that full information on the approval of the study protocol must also be provided in the manuscript.

## Human research participants

Policy information about studies involving human research participants

| Population characteristics | The covariate-relevant population characteristics of the human research participants of Memorial Sloan Kettering Cancer Center (MSKCC), Université Paris-Saclay, and the University of Manchester are provided below: Samples were obtained from acute myeloid leukemia (AML) patients treated at MSKCC, Université Paris-Saclay, and the University of Manchester. All samples were viably frozen and used to extract DNA, RNA, and protein. All subjects with AML were eligible for inclusion regardless of age, sex, or race. |
| --- | --- |
| Recruitment | All the participants were recruited without knowing their genotypes. Samples were genotyped and classified based on IDH2/SRSF2 genotypes. Samples that had mutations in IDH1, SF3B1, U2AF1, or ZRSR2 were excluded from RNA-seq and targeted RNA and protein analyses. Then RNA-seq was performed to analyze the splicing alterations. Therefore, there was no self-selection bias and it is unlikely that bias, if any, impacted the splicing analysis. |
| Ethics oversight | Studies were approved by the Institutional Review Boards of Memorial Sloan Kettering Cancer Center (under MSK IRB protocol 06-107), Université Paris-Saclay (under declaration DC-200-725 and authorization AC-2013-1884), and the University of Manchester (institution project approval 12-TISO-04), and conducted in accordance with the Declaration of Helsinki protocol. Written informed consent was obtained from all participants. Manchester samples were retrieved from the Manchester Cancer Research Centre Haematological Malignancy Tissue Biobank, which receives sample donations from all consenting leukemia patients presenting to The Christie Hospital (REC Reference 07/H1003/161+5; HTA license 30004; instituted with approval of the |

Note that full information on the approval of the study protocol must also be provided in the manuscript.

# ChIP-seq

## Data deposition

☒ Confirm that both raw and final processed data have been deposited in a public database such as GEO.

☒ Confirm that you have deposited or provided access to graph files (e.g. BED files) for the called peaks.

| | |
|---|---|
| Data access links<br>*May remain private before publication.* | The ChIP-seq data (and RNA-seq data  and eRRBS data) have been deposited in NCBI Sequence Read Archive (SRA) under accession number SRP133673. |
| Files in database submission | Sample_WT-WT-614, Sample_WT-WT-247, Sample_RQ-WT-584, Sample_RQ-WT-475, Sample_WT-PH-547, Sample_WT-PH-343, Sample_RQ-PH-524, Sample_RQ-PH-475, k562_WT_input, k562_WT_si_PolII-Ser2, k562_WT_si_PolII-Ser5, k562_p95_input, k562_p95H_si_PolII-Ser2, k562_p95H_si_PolII-Ser5 |
| Genome browser session<br>(e.g. UCSC) | The ChIP-seq data (and RNA-seq data  and eRRBS data) have been deposited in NCBI Sequence Read Archive (SRA) under accession number SRP133673. |

## Methodology

| | |
|---|---|
| Replicates | Two primary AML patient samples per IDH2/SRSF2 genotype (IDH2/SRSF2 WT/WT, Mutant/WT, WT/Mutant. and Mutant/Mutant) were used for the ChIP-seq experiments. |
| Sequencing depth | An average of 75 million paired reads was generated per sample (125 bp single-end). |
| Antibodies | Antibodies used for ChIP were as follows:<br>RNA polymerase II CTD repeat YSPTSPS (phospho S2) (Abcam; ab5095)<br>RNA polymerase II CTD repeat YSPTSPS (phospho S5) (Abcam; ab5408) |
| Peak calling parameters | Narrow peaks were called using the callpeak function from MACS2 v2.1.1.20160309 against matched input samples, using default parameters and a P-value cutoff of 1e-5, according to the ENCODE Histone ChIP-seq Data Standards and Processing Pipeline (https://www.encodeproject.org/chip-seq/histone/). |
| Data quality | For all samples, a P-value cutoff of 1e-5 against input was used. All peaks were called at a q-value of < 0.017. For each sample, the number of peaks with a fold-change > 5, and the average total number of peaks called is 19,200. |
| Software | ChIP-seq reads were mapped to the genome by calling Bowtie v1.0.048 with the arguments '-v 2 -k 1 -m 1 --best --strata'. Peaks were called using MACS2 v2.1.1.2016030952 against input control libraries with P < 1e-5 and subsequently filtered to remove peaks contained within ENCODE blacklisted regions and the mitochondrial genome. Subsequent data analysis was performed with Bioconductor in the R programming environment. Consensus peaks between samples were called using the soGGi package v1.14.0. Peaks were annotated using the ChIPseeker package v1.18.0. |

# Flow Cytometry

## Plots

Confirm that:

☒ The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).

☒ The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).

☒ All plots are contour plots with outliers or pseudocolor plots.

☒ A numerical value for number of cells or percentage (with statistics) is provided.

## Methodology

| | |
|---|---|
| Sample preparation | Surface-marker staining of hematopoietic cells was performed by first lysing cells with ACK lysis buffer and washing cells with ice-cold PBS. Cells were stained with antibodies in PBS/2% BSA for 30 minutes on ice. For hematopoietic stem/progenitor staining, cells were stained with the following antibodies: B220-APCCy7 (clone: RA3-6B2; purchased from BioLegend; catalog #: 103224; dilution: 1:200); B220-Bv711 (RA3-6B2; BioLegend, 103255; 1:200); CD3-PerCPCy5.5 (17A2; BioLegend; 100208; 1:200); CD3-APC (17A2; BioLegend; 100236; 1:200); CD3-APCCy7 (17A2; BioLegend; 100222; 1:200); Gr1-PECy7 (RB6-8C5; eBioscience; 25-5931-82; 1:500); CD11b-PE (M1/70; eBioscience; 12-0112-85; 1:500); CD11b-APCCy7 (M1/70; BioLegend; 101226; 1:200); CD11c-APCCy7 (N418; BioLegend; 117323; 1:200); NK1.1-APCCy7 (PK136; BioLegend; 108724; 1:200); Ter119-APCCy7 (BioLegend; 116223: 1:200); cKit-APC (2B8; BioLegend; 105812; 1:200); cKit-PerCPCy5.5  (2B8; BioLegend; 105824; 1:100); cKit-Bv605 (ACK2; BioLegend; 135120; 1:200); Sca1-PECy7 (D7; BioLegend; 108102; 1:200); CD16/CD32 (FcγRII/III)-Alexa700 (93; eBioscience; 56-0161-82; 1:200); CD34-FITC (RAM34; BD Biosciences; 553731; 1:200); CD45.1-FITC (A20; BioLegend; 110706; |

1:200); CD45.1-PerCPCy5.5 (A20; BioLegend; 110728; 1:200); CD45.1-PE (A20; BioLegend; 110708; 1:200); CD45.1-APC (A20; BioLegend; 110714; 1:200); CD45.2-PE (104; eBioscience; 12-0454-82; 1:200); CD45.2-Alexa700 (104; BioLegend; 109822; 1:200); CD45.2-Bv605 (104; BioLegend; 109841; 1:200); CD48-Bv711 (HM48-1; BioLegend; 103439; 1:200); CD150 (9D1; eBioscience; 12-1501-82; 1:200). DAPI was used to exclude dead cells. For sorting human leukemia cells, cells were stained with a lineage cocktail including CD34-PerCP (8G12; BD Biosciences; 345803; 1:200); CD117-PECy7 (104D2; eBioscience; 25-1178-42; 1:200); CD33-APC (P67.6; BioLegend; 366606; 1:200); HLA-DR-FITC (L243; BioLegend; 307604; 1:200); CD13-PE (L138; BD Biosciences; 347406; 1:200); CD45-APC-H7 (2D1; BD Biosciences; 560178; 1:200). The composition of mature hematopoietic cell lineages in the BM, spleen and peripheral blood was assessed using a combination of CD11b, Gr1, B220, and CD3. For the hematopoietic stem and progenitor analysis, a combination of CD11b, CD11c, Gr1, B220, CD3, NK1.1, and Ter119 was stained as lineage-positive cells.

| Instrument | All the FACS sorting was performed on FACS Aria, and analysis was performed on an LSRII or LSR Fortessa (BD Biosciences). |
|---|---|
| Software | FlowJo Ver.9 was used for analysis of flow cytometry data. |
| Cell population abundance | To check the purity of GFP and/or mCherry positivity in post-sort samples, the sorted samples were analyzed for GFP and/or mCherry by FACS Aria (BD Biosciences), and samples with >99% purity were used for analyses. |
| Gating strategy | The FSC/SSC gates of the starting cell population was set in order to include all the lineages of mouse hematopoietic cells such as granulocytes, monocytes, and lymphocytes. Then doublet cells were excluded by SSC-H vs SSC-W and FSC-H vs FSC-W gating. The boundaries between "positive" and "negative" staining cell population were defined by using unstained and single color-stained controls that were prepared by staining the whole BM mononuclear cells from B6 mice at 8-12 weeks with antibodies against mouse CD11b. "Positive" staining cell population was defined as CD11b+ population and "negative" staining cell population was defined by unstained control. The boundary for each fluorescence was set between these "positive" and "negative" staining cell populations. |

☒ Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.

# Structure of the inner kinetochore CCAN complex assembled onto a centromeric nucleosome

Kaige Yan[1,5], Jing Yang[1,5], Ziguo Zhang[1,5], Stephen H. McLaughlin[1], Leifu Chang[1,4], Domenico Fasci[2], Ann E. Ehrenhofer-Murray[3], Albert J. R. Heck[2] & David Barford[1]*

In eukaryotes, accurate chromosome segregation in mitosis and meiosis maintains genome stability and prevents aneuploidy. Kinetochores are large protein complexes that, by assembling onto specialized Cenp-A nucleosomes[1,2], function to connect centromeric chromatin to microtubules of the mitotic spindle[3,4]. Whereas the centromeres of vertebrate chromosomes comprise millions of DNA base pairs and attach to multiple microtubules, the simple point centromeres of budding yeast are connected to individual microtubules[5,6]. All 16 budding yeast chromosomes assemble complete kinetochores using a single Cenp-A nucleosome (Cenp-A$^{Nuc}$), each of which is perfectly centred on its cognate centromere[7–9]. The inner and outer kinetochore modules are responsible for interacting with centromeric chromatin and microtubules, respectively. Here we describe the cryo-electron microscopy structure of the *Saccharomyces cerevisiae* inner kinetochore module, the constitutive centromere associated network (CCAN) complex, assembled onto a Cenp-A nucleosome (CCAN–Cenp-A$^{Nuc}$). The structure explains the interdependency of the constituent subcomplexes of CCAN and shows how the Y-shaped opening of CCAN accommodates Cenp-A$^{Nuc}$ to enable specific CCAN subunits to contact the nucleosomal DNA and histone subunits. Interactions with the unwrapped DNA duplex at the two termini of Cenp-A$^{Nuc}$ are mediated predominantly by a DNA-binding groove in the Cenp-L–Cenp-N subcomplex. Disruption of these interactions impairs assembly of CCAN onto Cenp-A$^{Nuc}$. Our data indicate a mechanism of Cenp-A nucleosome recognition by CCAN and how CCAN acts as a platform for assembly of the outer kinetochore to link centromeres to the mitotic spindle for chromosome segregation.

The 14-subunit CCAN complex assembled onto specialized Cenp-A nucleosomes (in which Cenp-A is substituted for histone H3) reconstituted using either an *S. cerevisiae* centromere sequence or the Widom 601 sequence, with both complexes eluting at similar volumes on size-exclusion chromatography (SEC) (Extended Data Fig. 1a–e). By contrast, CCAN did not assemble onto a canonical H3 nucleosome, indicating the specificity of the CCAN–Cenp-A$^{Nuc}$ interaction (Extended Data Fig. 1b, f). Cryo-electron microscopy (cryo-EM) of CCAN–Cenp-A$^{Nuc}$ (using the more stable Widom 601–Cenp-A$^{Nuc}$) revealed a heterogeneous population of particles that, by 3D classification, were identified as monomeric free CCAN, a monomer of CCAN in complex with Cenp-A$^{Nuc}$ and dimeric CCAN (Extended Data Figs. 2, 3). A 3D reconstruction of free monomeric CCAN was determined to 3.5 Å resolution (Fig. 1, Extended Data Figs. 2, 3, Extended Data Table 1). Clearly defined electron microscopy density for the majority of amino acid side chains (Extended Data Fig. 4, Extended Data Tables 1, 2) enabled building and refinement of the complete atomic model of CCAN, guided by existing models of individual CCAN subunits. The CCAN–Cenp-A$^{Nuc}$ complex at 4.15 Å was built by docking apo-CCAN and a nucleosome into the CCAN–Cenp-A$^{Nuc}$ cryo-EM reconstruction (Fig. 2, Extended Data Table 1). A cryo-EM reconstruction

of uncrosslinked CCAN–Cenp-A$^{Nuc}$, at lower resolution (Extended Data Fig. 5a), matched that of the crosslinked structure, whereas the reconstruction of the free CCAN dimer, determined at 8.6 Å (Extended Data Fig. 5b), resembles the 4.25 Å structure of *S. cerevisiae* CCAN[10], although the Nkp1 and Nkp2 subunit assignments differ.

The arrangement of the three subcomplexes of CCAN; Cenp-L–Cenp-N (hereafter Cenp-LN), Cenp-O–Cenp-P–Cenp-Q–Cenp-U–Nkp1–Nkp2 (Cenp-OPQU+) and Cenp-H–Cenp-I–Cenp-K–(Cenp-T–Cenp-W) (Cenp-HIK–TW) (Extended Data Table 2), generates a Y-shaped structure (Fig. 1a, b). The Cenp-N subunit, located at the centre of the Y-shaped structure, is the coordinating element of CCAN, consistent with it forming a critical node at the centromere–kinetochore interface[11]. Cenp-OPQU+, which has an elongated shape and generates the stem and one arm of the Y, interacts mainly with Cenp-N. Cenp-L also forms an extensive interface with Cenp-N, and contributes the major point of contact with Cenp-HIK–TW. Together, Cenp-L and Cenp-HIK–TW form the opposite arm of the Y (Fig. 1a, b). The six-subunit Cenp-OPQU+ module shares four subunits in common with vertebrate Cenp-OPQUR, and its structure in CCAN resembles the negative-stain electron microscopy reconstruction of human CENP-OPQUR[12]. The long N-terminal regions of Cenp-O and Cenp-P, which are disordered in the *Kluyveromyces lactis* crystal structure[13], are more structured through interactions with Cenp-HIK and Cenp-N (Fig. 1b, c). Four subunits of Cenp-OPQU+ (Cenp-Q, Cenp-U, Nkp1 and Nkp2) form extended α-helices that associate in a parallel, interweaved fashion to create an irregular coiled-coil α-helical bundle. This shares a marked similarity to the outer kinetochore complex Mis12[14,15] (Extended Data Fig. 5c). Nkp1 and Nkp2 create an outer layer of α-helices in Cenp-OPQU+, which are probably substituted by Cenp-R in vertebrates[12].

The Cenp-HIK module (Fig. 1c), which resembles the free Cenp-HIK complex (Extended Data Fig. 5d), is dominated by the C-terminal HEAT motif repeats of Cenp-I (Extended Data Fig. 4e). The coiled-coil α-helices of Cenp-H and Cenp-K run anti-parallel to Cenp-I (Fig. 1c, Extended Data Fig. 4a–c). The base of Cenp-HIK is a four α-helical bundle comprising the N termini of Cenp-H and Cenp-K. The flexible head domain, present in free Cenp-HIK (Cenp-HIK$^{Head}$), and a small population of CCAN particles (Extended Data Figs. 3c, 5b, d), matches the shape of the crystal structure of the N-terminal Cenp-I HEAT repeats that are associated with the C termini of both Cenp-H and Cenp-K[16] (Fig. 1d). The Cenp-TW subcomplex, comprising the histone-fold domain (HFD) subunits Cenp-T and Cenp-W, is not clearly resolved in cryo-EM maps of CCAN and CCAN–Cenp-A$^{Nuc}$. Cenp-TW associates with Cenp-HIK in solution, consistent with previous studies[11,17], and the HFD domains of Cenp-T and Cenp-W (Cenp-T$^{HFD}$W) interacts equally well with a complex comprising Cenp-HIK$^{Head}$ (Extended Data Fig. 1g–j), indicating that the HFDs of Cenp-TW interact directly with Cenp-HIK$^{Head}$.

The relative organization of CCAN subunits in our cryo-EM reconstruction is in agreement with that defined from the de novo assembly
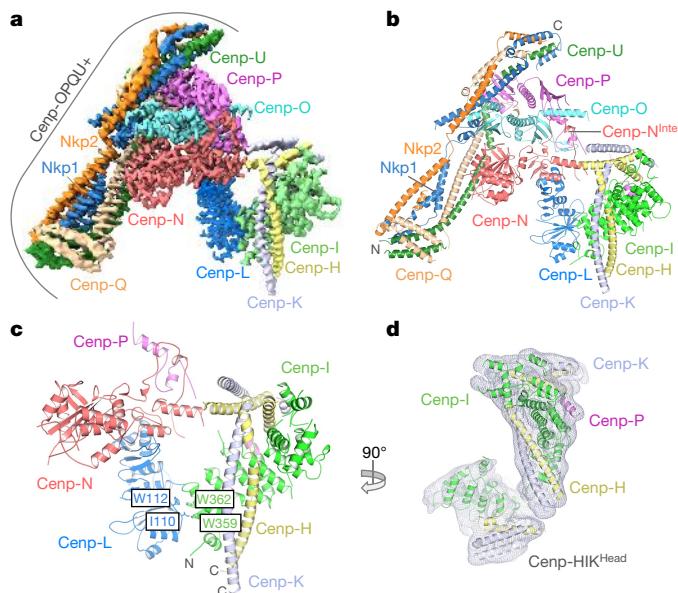
**Fig. 1 | Structure of the *S. cerevisiae* CCAN complex. a**, **b**, Cryo-EM density map (**a**) and cartoon representation of CCAN (**b**). Eleven subunits are assigned. N and C indicate the N and C termini of Cenp-QU, Nkp1 and Nkp2. **c**, Details of the Cenp-HIK–Cenp-LN interface. Residues of Cenp-I are visible from residue 320 onwards. **d**, Cryo-EM density for the complete Cenp-HIK module showing Cenp-HIK$^{Head}$ from the CCAN dimer cryo-EM 3D class (Extended Data Figs. 3a, 5b).

of the *S. cerevisiae* kinetochore[9] (Extended Data Fig. 1k) and consistent with a negative-stain electron microscopy reconstruction of the human CENP-HIKM–LN–OPQUR complex[12]. To assess the validity of our structure, we performed crosslinking mass spectrometry (XL-MS) analysis of the complexes. Numerous intra- and intersubunit crosslinks were identified (Extended Data Fig. 6a, b, Supplementary Tables 1, 2). Mapping these crosslinks onto CCAN and CCAN–Cenp-A$^{Nuc}$, for

which both lysines of the crosslinked pair are defined, showed that 95% of the detected crosslinks are within the expected linker-distance constraints (Extended Data Fig. 6c–f).

Kinetochores assemble onto Cenp-A$^{Nuc}$ (refs [9,18]), the hallmark of centromeric chromatin, with the CCAN subunits Cenp-C and Cenp-N directing this assembly[19,20]. In the CCAN–Cenp-A$^{Nuc}$ complex, Cenp-A$^{Nuc}$ is an octameric nucleosome, with DNA wrapped as a left-handed superhelix (Fig. 2, Supplementary Video 1), as previously shown for free Cenp-A$^{Nuc}$ (refs [8,21–23]). Consistent with these reports is that compared with canonical H3 nucleosomes, in the CCAN–Cenp-A$^{Nuc}$ complex, the DNA gyre of Cenp-A$^{Nuc}$ is more loosely wrapped. In CCAN–Cenp-A$^{Nuc}$, only 105 bp of DNA encircle the Cenp-A-octamer, compared with 147 bp for canonical nucleosomes[24] (Figs. 2, 3a, b). A total of 20 bp of DNA are unwrapped equally at each DNA terminus of Cenp-A$^{Nuc}$. One of the unwrapped DNA termini, well defined in cryo-EM density, interacts with CCAN, whereas the other is disordered (Fig. 2a). We observe clearly defined α-helical density for the N-terminal segment of one Cenp-A subunit (Cenp-A$^N$), which is inserted between the unwrapped DNA duplex and DNA gyre (Figs. 2a, 3c).

In the CCAN–Cenp-A$^{Nuc}$ complex (Fig. 2, Supplementary Video 1), Cenp-A$^{Nuc}$ inserts end-on into the Y-shaped opening of CCAN, with each arm of CCAN embracing opposite sides of the nucleosome. This positions the Cenp-LN module to form extensive contacts with the unwrapped DNA duplex at one of the termini of the Cenp-A$^{Nuc}$ DNA gyre (Fig. 2). Cenp-LN adopts a U-shaped structure, creating an evolutionarily conserved, positively charged groove that engages the unwrapped DNA (Fig. 3b, Extended Data Fig. 7a–c). The DNA duplex runs along the Cenp-LN groove, exiting opposite to the nucleosome (Figs. 2, 3a, b). Cenp-HIK$^{Head}$–Cenp-TW also functions in Cenp-A$^{Nuc}$ recognition, as indicated by the CCAN–Cenp-A$^{Nuc}$ complex, in which cryo-EM density corresponding to Cenp-HIK$^{Head}$–Cenp-TW contacts the DNA gyre of Cenp-A$^{Nuc}$, with Cenp-I in close proximity to Cenp-A (Fig. 2c, Extended Data Fig. 3c, Supplementary Video 1). Compared with apo-CCAN, Cenp-HIK$^{Head}$–Cenp-TW rotates by around 90° to accommodate Cenp-A$^{Nuc}$ (Extended Data Fig. 5e). Previous studies have suggested that the vertebrate Cenp-TWSX heterotetramer forms
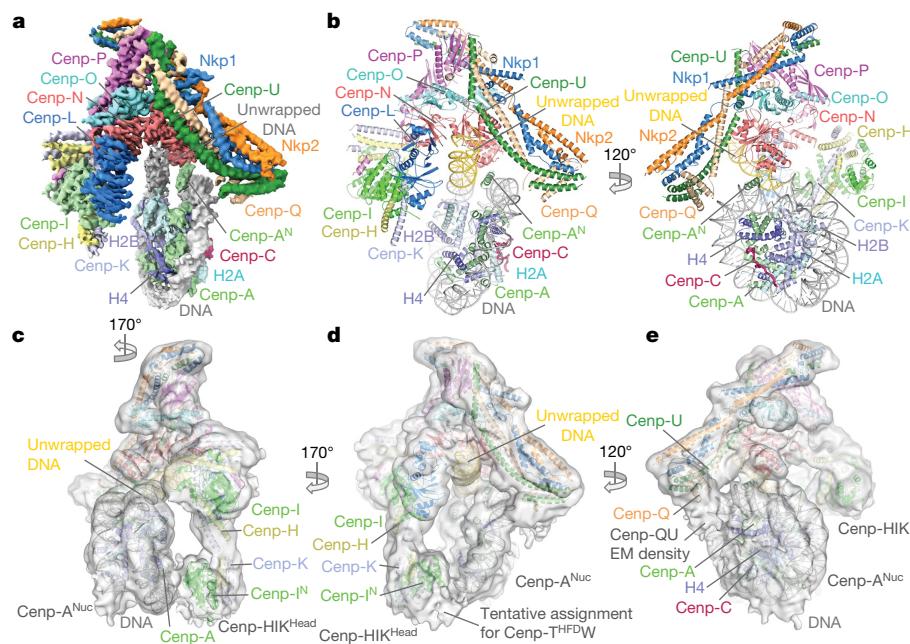


**Fig. 2 | Structure of the *S. cerevisiae* CCAN–Cenp-A$^{Nuc}$ complex. a**, Cryo-EM density map of CCAN–Cenp-A$^{Nuc}$. Cenp-A$^N$ comprises residues 111–129. **b**, Two views of a cartoon representation of CCAN–Cenp-A$^{Nuc}$. Cenp-A$^{Nuc}$ wraps about 105 bp of DNA, leaving 20 bp of DNA unwrapped at both ends (coloured yellow for the ordered terminal segment; Supplementary Video 1). **c–e**, Three views of the cryo-EM

density of a 3D subclass of the overall CCAN–Cenp-A$^{Nuc}$ 3D class, before application of the mask used to refine the cryo-EM map shown in **a** (Extended Data Fig. 3a), highlighting contacts to Cenp-A$^{Nuc}$. **c**, The Cenp-HIK$^{Head}$ module contacts Cenp-A. **d**, Cenp-T$^{HFD}$W contacts the DNA gyre of Cenp-A$^{Nuc}$. **e**, The N-terminal region of Cenp-QU contacts Cenp-A and H4.
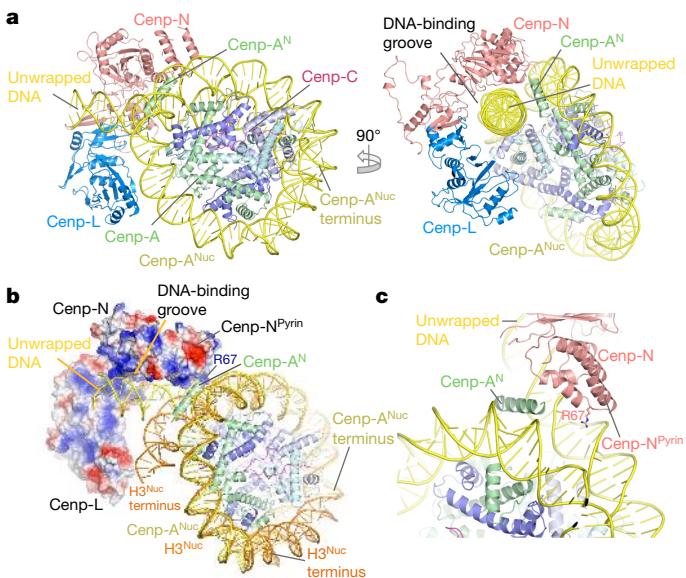
**Fig. 3 | Cenp-LN interacts with the unwrapped DNA duplex of Cenp-A$^{Nuc}$.** **a**, Two orthogonal views showing the unwrapped DNA duplex of Cenp-A$^{Nuc}$ engaged by the DNA-binding groove of the Cenp-LN subcomplex. **b**, Surface of Cenp-LN showing positive electrostatic potential of the DNA-binding groove. The canonical *S. cerevisiae* H3 nucleosome (orange, Protein Data Bank (PDB) ID: 1ID3[24]) wraps 147 bp of DNA compared with the 105 bp wrapped by the *S. cerevisiae* Cenp-A nucleosome (yellow). **c**, Magnified view showing insertion of the N terminus of Cenp-A (Cenp-A$^N$) between the unwrapped DNA duplex and DNA gyre of Cenp-A$^{Nuc}$. Arg67 of the Cenp-N pyrin domain inserts into the DNA major groove.

a nucleosome-like particle to interact with DNA[25]. However, this is not compatible with *S. cerevisiae* Cenp-TW exactly co-localizing with centromeric Cenp-A$^{Nuc}$ in a Cenp-I-dependent manner[17]. The HFDs of Cenp-TW were assigned to cryo-EM density associated with Cenp-HIK$^{Head}$ contacting the DNA gyre of Cenp-A$^{Nuc}$, visible in a minor 3D class of CCAN–Cenp-A$^{Nuc}$ (Fig. 2d, Extended Data Fig. 3c). On the opposite side of CCAN to Cenp-HIK, the N-terminal regions of Cenp-Q and Cenp-U contact the DNA gyre of Cenp-A$^{Nuc}$ and the N termini of Cenp-A and H4 (Fig. 2b, right, e). This is consistent with the Cenp-Q–Cenp-U (Cenp-QU) dimer binding DNA[26] and recognizing the posttranslational status of the N terminus of Cenp-A[27], and further validated by our XL-MS data revealing Cenp-Q crosslinks to H2A and H2B (Extended Data Fig. 6b).

Cenp-N engages Cenp-A$^{Nuc}$ in the budding yeast CCAN–Cenp-A$^{Nuc}$ complex in a different manner to how isolated vertebrate Cenp-N subunit interacts with Cenp-A$^{Nuc}$ through the L1 loop of Cenp-A and the adjacent DNA gyre[28,29]. Because of steric clashes, the interaction of Cenp-N with Cenp-A$^{Nuc}$ revealed in these studies is not compatible with the position of Cenp-N in the context of the CCAN complex (Extended Data Fig. 7d). Binding of Cenp-A$^{Nuc}$ at this interface of CCAN, as previously proposed[10], would require substantial conformational changes of CCAN. The discrepancy between our structure and that of the vertebrate system may either reflect genuine species differences in CCAN–Cenp-A$^{Nuc}$ architectures or result from the vertebrate Cenp-N–Cenp-A$^{Nuc}$ structure representing an intermediate in the CCAN–Cenp-A$^{Nuc}$ assembly pathway, in accordance with CCAN–Cenp-A$^{Nuc}$ remodelling during the cell cycle[11].

Cenp-C also determines kinetochore–Cenp-A$^{Nuc}$ interactions[20], and we found that Cenp-C is required for stable assembly onto Cenp-A-*Cen3* nucleosomes (data not shown), although not Cenp-A–Widom 601 nucleosomes (Fig. 4b). Cenp-C interacts with Cenp-A through its Cenp-C motif (Extended Data Fig. 5f), similar to vertebrates[30]. However, the regions of Cenp-C associated with CCAN were not visible in the cryo-EM maps. XL-MS data indicate that Cenp-C participates

in multiple interactions with CCAN (Extended Data Fig. 6a, b, g, Supplementary Tables 1, 2).

To test the validity of the CCAN–Cenp-A$^{Nuc}$ structure, we mutated 13 Arg and Lys residues in Cenp-N that line the Cenp-LN DNA-binding groove (Fig. 4a) and tested the ability of the mutant CCAN to assemble onto Cenp-A$^{Nuc}$. To avoid complications of Cenp-C interacting with Cenp-A$^{Nuc}$, we used CCAN without Cenp-C (CCAN$^{\Delta Cenp-C}$). The Cenp-N mutant (Cenp-N$^{Mut}$) did not impair CCAN$^{\Delta Cenp-C}$ assembly, and similar to CCAN, CCAN$^{\Delta Cenp-C}$ binds to Cenp-A–Widom 601 nucleosomes, but not H3 nucleosomes (Fig. 4b, Extended Data Figs. 8a–c, 9a, b). Cenp-N$^{Mut}$ disrupted CCAN$^{\Delta Cenp-C}$–Cenp-A$^{Nuc}$ interactions (Fig. 4b, Extended Data Fig. 8d). By contrast, mutating the L1 loop of Cenp-A did not disrupt the binding of CCAN$^{\Delta Cenp-C}$ to Cenp-A$^{Nuc}$ (Extended Data Figs. 8e, 9a).

We then assessed the role of the unwrapped DNA termini of Cenp-A$^{Nuc}$ in mediating CCAN–Cenp-A$^{Nuc}$ interactions. Because the αN-helix of the H3 histone stabilizes the wrapped DNA termini of canonical H3 nucleosomes[22,24], to create a more closed, highly wrapped Cenp-A$^{Nuc}$, we substituted the N-terminal 50 residues of H3 for the N-terminal 140 residues of Cenp-A, creating a chimeric H3$^N$–Cenp-A (Extended Data Fig. 7e–g). The resultant H3$^N$–Cenp-A$^{Nuc}$ wrapped a similar length of DNA as did H3$^{Nuc}$ (approximately 147 bp) (Extended Data Fig. 9c). The affinity of CCAN$^{\Delta Cenp-C}$ for H3$^N$–Cenp-A$^{Nuc}$ was severely disrupted, such that CCAN$^{\Delta Cenp-C}$ was substantially dissociated from H3$^N$–Cenp-A$^{Nuc}$ (Fig. 4b, Extended Data Fig. 8g). Binding of H3$^N$-Cenp-A$^{Nuc}$ to CCAN$^{\Delta Cenp-C}$ was completely disrupted with Cenp-N$^{Mut}$ (Fig. 4b, Extended Data Fig 8h). The reduced affinity of CCAN for H3$^N$–Cenp-A$^{Nuc}$ is not due to the lack of the Cenp-A N terminus, because CCAN bound to full-length Cenp-A$^{Nuc}$ and Cenp-A$^{Nuc}$ in which residues 1–129 of Cenp-A are deleted ($^{\Delta N}$Cenp-A$^{Nuc}$) equally well (Fig. 4b, Extended Data Fig. 8c, f). These biochemical studies confirm that CCAN interacts with the unwrapped DNA termini of Cenp-A$^{Nuc}$ and that a major role of the Cenp-LN DNA-binding groove is to engage the unwrapped DNA gyre of Cenp-A$^{Nuc}$, as shown by the CCAN–Cenp-A$^{Nuc}$ cryo-EM structure (Fig. 3b).

Disruption of the *S. cerevisiae* Cenp-N gene (*CHL4*) causes chromosome loss and instability without affecting viability[31]. However, combining a *chl4* deletion with either mutation of Cenp-A (*CSE4*) or deletion of other kinetochore subunits results in synthetic growth defects and lethality[9,27]. Cenp-N is an essential gene in *Schizosaccharomyces pombe* and humans. To investigate the in vivo consequences of disrupting the DNA-binding groove of Cenp-LN, we tested whether the synthetic growth defect of the *chl4Δ cse4-R37A* mutant at 37 °C in *S. cerevisiae* (ref. [27]) was rescued by Cenp-N$^{Mut}$. Whereas wild type Cenp-N rescued the growth defect of the *chl4Δ cse4-R37A* mutant, Cenp-N$^{Mut}$ did not (Fig. 4c–d). This result demonstrates a functional role for the Cenp-LN DNA-binding groove, and together with our biochemical data (Fig. 4b, Extended Data Fig. 8), supports the CCAN–Cenp-A$^{Nuc}$ architecture that we report here. In *S. cerevisiae*, Cenp-A$^{Nuc}$ is linked to the outer kinetochore Ndc80 complex and associated microtubules through a pathway comprising the essential proteins Cenp-C, Cenp-QU and the Mis12 complex and by a second pathway involving Cenp-TW and Cenp-N[9] (Extended Data Fig. 1k). The location of Cenp-N at the centre of CCAN is consistent with these two pathways. The unwrapped DNA termini of Cenp-A$^{Nuc}$ contribute to stabilizing the CCAN–Cenp-A$^{Nuc}$ complex through the Cenp-LN DNA-binding groove, augmented by contacts of both Cenp-A and the Cenp-A$^{Nuc}$ DNA gyre with Cenp-C (Extended Data Fig. 5f), Cenp-LN (Fig. 3b), Cenp-TW, Cenp-HIK$^{Head}$ and Cenp-QU[27] (Fig. 2c–e).

In the cryo-EM reconstruction, Cenp-A$^{Nuc}$ is associated with a single CCAN, whereas the expected stoichiometry is two CCANs to one Cenp-A$^{Nuc}$ (ref. [32]). SEC with multi-angle light scattering (SEC–MALS) and analytical ultracentrifugation confirmed that the reconstituted CCAN–Cenp-A$^{Nuc}$ is consistent with two CCANs per Cenp-A$^{Nuc}$ ((CCAN)$_2$–Cenp-A$^{Nuc}$ complex) (Extended Data Fig. 10a–g). In a generated model of (CCAN)$_2$–Cenp-A$^{Nuc}$, two CCAN complexes associate through their tips of the Y, creating a slot that perfectly accommodates
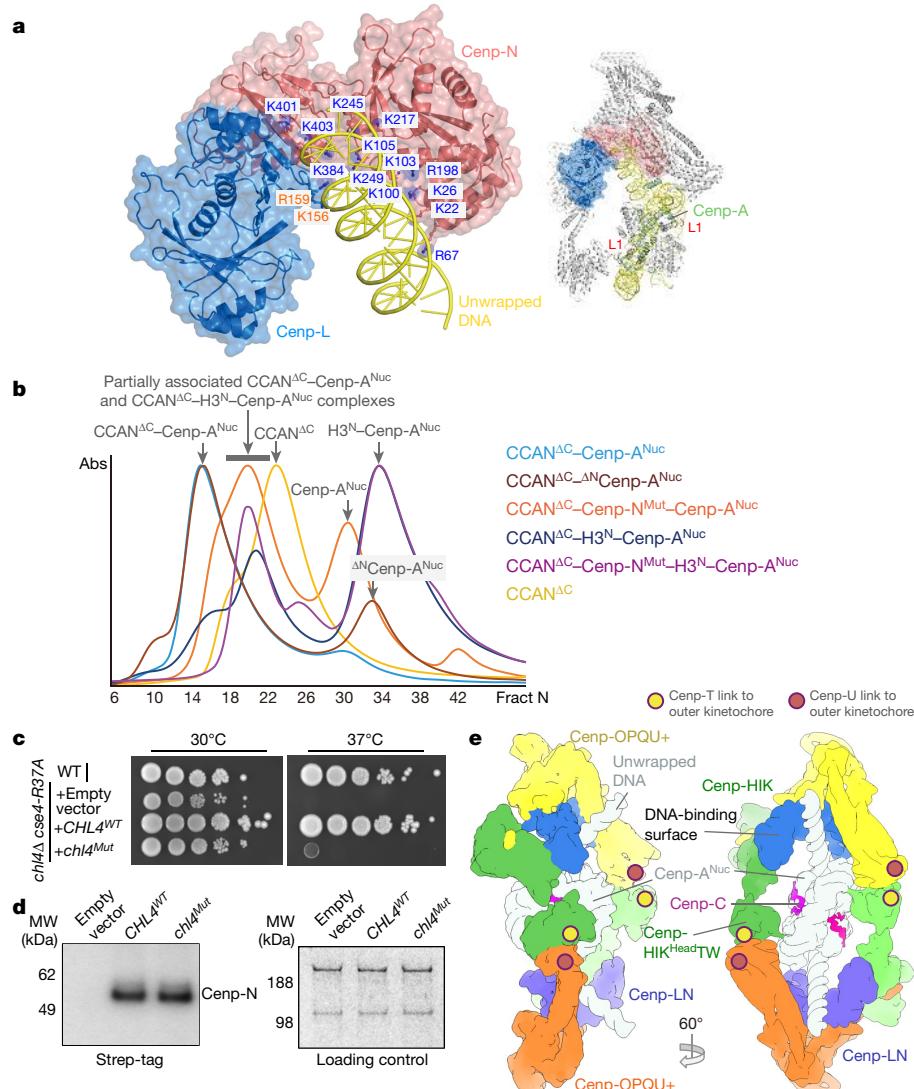
**Fig. 4 | The Cenp-N DNA-binding groove is required for stable CCAN–Cenp-A$^{Nuc}$ interactions. a**, Left, surface of the Cenp-LN module showing the Cenp-N DNA-binding groove engaging the unwrapped DNA, indicating the 13 mutated Arg and Lys residues of Cenp-N (blue labels). Right, overview of CCAN–Cenp-A$^{Nuc}$ showing the Cenp-A L1 loop. **b**, Size-exclusion chromatograms of various CCAN$^{\Delta Cenp-C}$–Cenp-A$^{Nuc}$ complexes. Wild-type CCAN$^{\Delta Cenp-C}$ forms a complex with Cenp-A$^{Nuc}$, but mutating the Cenp-N DNA-binding groove weakens CCAN–Cenp-A$^{Nuc}$ interactions (Extended Data Fig. 8c, d). The binding of both CCAN$^{\Delta Cenp-C}$ and CCAN$^{\Delta Cenp-C}$–Cenp-N$^{Mut}$ to H3$^N$–Cenp-A$^{Nuc}$ is severely disrupted, and few complexes formed (Extended Data Fig. 8g, h). The positions of complexes are indicated by arrows. (CCAN$^{\Delta C}$ refers to CCAN$^{\Delta Cenp-C}$). This experiment was performed independently in triplicate with similar

results. **c**, The DNA-binding groove functions in vivo in *S. cerevisiae*. Wild-type Cenp-N ($CHL4^{WT}$) rescues the growth defect of the *chl4Δ cse4-R73A* mutant strain at 37 °C, whereas the Cenp-N$^{Mut}$ (*chl4$^{Mut}$*) does not. WT, wild-type strain. This experiment was performed independently ten times with similar results. **d**, Western blot showing that Cenp-N$^{WT}$ and Cenp-N$^{Mut}$ are expressed at equivalent levels in the *chl4Δ cse4-R73A* mutant strain (left) and loading control (right; Coomassie-blue-stained gel shows dynein and acetyl-CoA carboxylase). Experiments in **d** were performed independently in triplicate with similar results. **e**, Two views showing a representation of the (CCAN)$_2$–Cenp-A$^{Nuc}$ complex with the second CCAN protomer generated by the dyad symmetry of Cenp-A$^{Nuc}$. Sites of contact to the outer kinetochore (through Cenp-U and Cenp-T) are indicated. For gel source data, see Supplementary Fig. 1.

Cenp-A$^{Nuc}$ that is inserted vertically (Fig. 4e). The two CCAN complexes cradle Cenp-A$^{Nuc}$ with its unwrapped DNA duplexes stretched out, overlying the DNA-binding surface of CCAN, consistent with XL-MS crosslinks between Cenp-Q and Cenp-TW (Extended Data Fig. 6b). Extensive 2D classification of the cryo-EM data identified 2D classes of (CCAN)$_2$–Cenp-A$^{Nuc}$ particles with two-fold symmetry axes (Extended Data Fig. 2c). These particles correspond closely to the calculated reprojections of the proposed (CCAN)$_2$–Cenp-A$^{Nuc}$ complex (Extended Data Fig. 10h). Cryo-EM grids destabilize CCAN–Cenp-A$^{Nuc}$, resulting in a very low abundance of (CCAN)$_2$–Cenp-A$^{Nuc}$ particles.

In *S. cerevisiae*, the CBF3 complex engages the CDEIII element of the approximately 125-bp centromere to direct Cenp-A nucleosome deposition. Modelling indicates that Cenp-A$^{Nuc}$ can simultaneously

accommodate CBF3 only when bound to a single CCAN promoter (Extended Data Fig. 9d), which suggests that CBF3 would not associate with a fully assembled kinetochore.

The (CCAN)$_2$–Cenp-A$^{Nuc}$ model suggests two possibilities for how a kinetochore-attached microtubule would segregate centromeric chromatin (Extended Data Fig. 10i, j, Supplementary Video 2). In one scenario, CCAN attaches to the microtubule through the outer kinetochore using the same face as its DNA-binding surface (Extended Data Fig. 10i). This would sandwich the DNA between CCAN and the outer kinetochore, a possibility compatible with the long flexible linkers that attach CCAN to the outer kinetochore. As the microtubule pulls on the kinetochore, CCAN would hoist the overlying DNA. Alternatively, microtubules could attach to CCAN from the opposite face to its DNA-binding surface, so the chromosome is pulled from behind the inner

kinetochore (Extended Data Fig. 10j). Because vertebrate Cenp-A$^{Nuc}$ also wraps between 100–120 bp (of α-satellite DNA)[22] with nucleosome unwrapping enhanced by Cenp-C[33] and the human CCAN architecture[12] is similar to that of yeast, it is likely that the mechanism of recognition of the specialized Cenp-A nucleosome that we describe here for the budding yeast inner kinetochore is evolutionarily conserved.

## Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-019-1609-1.

1. Earnshaw, W. C. & Rothfield, N. Identification of a family of human centromere proteins using autoimmune sera from patients with scleroderma. *Chromosoma* **91**, 313–321 (1985).
2. Meluh, P. B., Yang, P., Glowczewski, L., Koshland, D. & Smith, M. M. Cse4p is a component of the core centromere of *Saccharomyces cerevisiae*. *Cell* **94**, 607–613 (1998).
3. Cheeseman, I. M. The kinetochore. *Cold Spring Harb. Perspect. Biol.* **6**, a015826 (2014).
4. Musacchio, A. & Desai, A. A molecular view of kinetochore assembly and function. *Biology* **6**, 5 (2017).
5. Clarke, L. & Carbon, J. Isolation of a yeast centromere and construction of functional small circular chromosomes. *Nature* **287**, 504–509 (1980).
6. Winey, M. et al. Three-dimensional ultrastructural analysis of the *Saccharomyces cerevisiae* mitotic spindle. *J. Cell Biol.* **129**, 1601–1615 (1995).
7. Furuyama, S. & Biggins, S. Centromere identity is specified by a single centromeric nucleosome in budding yeast. *Proc. Natl Acad. Sci. USA* **104**, 14706–14711 (2007).
8. Camahort, R. et al. Cse4 is part of an octameric nucleosome in budding yeast. *Mol. Cell* **35**, 794–805 (2009).
9. Lang, J., Barber, A. & Biggins, S. An assay for de novo kinetochore assembly reveals a key role for the CENP-T pathway in budding yeast. *eLife* **7**, e37819 (2018).
10. Hinshaw, S. M. & Harrison, S. C. The structure of the Ctf19c/CCAN from budding yeast. *eLife* **8**, e44239 (2019).
11. McKinley, K. L. et al. The CENP-L–N complex forms a critical node in an integrated meshwork of interactions at the centromere-kinetochore interface. *Mol. Cell* **60**, 886–898 (2015).
12. Pesenti, M. E. et al. Reconstitution of a 26-subunit human kinetochore reveals cooperative microtubule binding by CENP-OPQUR and NDC80. *Molecular Cell* **71**, 923–939 (2018).
13. Schmitzberger, F. & Harrison, S. C. RWD domain: a recurring module in kinetochore architecture shown by a Ctf19–Mcm21 complex structure. *EMBO Rep.* **13**, 216–222 (2012).
14. Dimitrova, Y. N., Jenni, S., Valverde, R., Khin, Y. & Harrison, S. C. Structure of the MIND complex defines a regulatory focus for yeast kinetochore assembly. *Cell* **167**, 1014–1027 (2016).
15. Petrovic, A. et al. Structure of the MIS12 complex and molecular basis of its interaction with CENP-C at human kinetochores. *Cell* **167**, 1028–1040 (2016).
16. Hu, L. et al. Structural analysis of fungal CENP-H/I/K homologs reveals a conserved assembly mechanism underlying proper chromosome alignment. *Nucleic Acids Res.* **47**, 468–479 (2019).
17. Pekgöz Altunkaya, G. et al. CCAN assembly configures composite binding interfaces to promote cross-linking of Ndc80 complexes at the kinetochore. *Curr. Biol.* **26**, 2370–2378 (2016).
18. Guse, A., Carroll, C. W., Moree, B., Fuller, C. J. & Straight, A. F. In vitro centromere and kinetochore assembly on defined chromatin templates. *Nature* **477**, 354–358 (2011).
19. Carroll, C. W., Silva, M. C., Godek, K. M., Jansen, L. E. & Straight, A. F. Centromere assembly requires the direct recognition of CENP-A nucleosomes by CENP-N. *Nat. Cell Biol.* **11**, 896–902 (2009).
20. Carroll, C. W., Milks, K. J. & Straight, A. F. Dual recognition of CENP-A nucleosomes is required for centromere assembly. *J. Cell Biol.* **189**, 1143–1155 (2010).
21. Kingston, I. J., Yung, J. S. & Singleton, M. R. Biophysical characterization of the centromere-specific nucleosome from budding yeast. *J. Biol. Chem.* **286**, 4021–4026 (2011).
22. Tachiwana, H. et al. Crystal structure of the human centromeric nucleosome containing CENP-A. *Nature* **476**, 232–235 (2011).
23. Roulland, Y. et al. The flexible ends of CENP-A nucleosome are required for mitotic fidelity. *Mol. Cell* **63**, 674–685 (2016).
24. White, C. L., Suto, R. K. & Luger, K. Structure of the yeast nucleosome core particle reveals fundamental changes in internucleosome interactions. *EMBO J.* **20**, 5207–5218 (2001).
25. Nishino, T. et al. CENP-T–W–S–X forms a unique centromeric chromatin structure with a histone-like fold. *Cell* **148**, 487–501 (2012).
26. Hornung, P. et al. A cooperative mechanism drives budding yeast kinetochore assembly downstream of CENP-A. *J. Cell Biol.* **206**, 509–524 (2014).
27. Anedchenko, E. A. et al. The kinetochore module Okp1$^{CENP-Q}$/Ame1$^{CENP-U}$ is a reader for N-terminal modifications on the centromeric histone Cse4$^{CENP-A}$. *EMBO J.* **38**, e98991 (2019).
28. Chittori, S. et al. Structural mechanisms of centromeric nucleosome recognition by the kinetochore protein CENP-N. *Science* **359**, 339–343 (2018).
29. Pentakota, S. et al. Decoding the centromeric nucleosome through CENP-N. *eLife* **6**, e33442 (2017).
30. Kato, H. et al. A conserved mechanism for centromeric nucleosome recognition by centromere protein CENP-C. *Science* **340**, 1110–1113 (2013).
31. Kouprina, N. et al. Identification and cloning of the CHL4 gene controlling chromosome segregation in yeast. *Genetics* **135**, 327–341 (1993).
32. Weir, J. R. et al. Insights from biochemical reconstitution into the architecture of human kinetochores. *Nature* **537**, 249–253 (2016).
33. Falk, S. J. et al. Chromosomes. CENP-C reshapes and stabilizes CENP-A nucleosomes at the centromere. *Science* **348**, 699–703 (2015).

## METHODS

No statistical methods were used to predetermine sample size. The experiments were not randomized. The investigators were not blinded to allocation during experiments and outcome assessment.

**Cloning, expression, purification and reconstitution of recombinant CCAN–Cenp-A$^{Nuc}$ nucleosome complex.** *Cloning.* The genes for *CTF19*, *OKP1*, *MCM21*, *AME1*, *NKP1*, *NKP2*, *CTF3*, *MCM16*, *MCM22*, *CNN1*, *WIP1*, *MIF2*, *CHL4* and *IML3* (*MCM19*) (see Extended Data Table 2 for vertebrate Cenp homologues) were amplified by PCR from *S. cerevisiae* genomic DNA and cloned into a pU1 plasmid using a modified Multibac expression system[34]. The intron in *MCM21* was deleted by the USER method[34]. A double StrepII tag together with a TEV cleavage site was attached to the C termini of Ame1, Ctf3, Chl4, Mif2 and Cnn1 proteins. For expression of the Cenp-OPUQ+ complex (also called COMA+: Ctf19, Okp1, Mcm21, Ame1, Nkp1 and Nkp2 gene expression cassettes in pU1 were subsequently cloned into a pF2 vector[34]. The gene expression cassettes for *CTF3*, *MCM16*, *MCM22*, *CNN1* and *WIP1* were cloned into pF2 to generate the Cenp-HIK–TW complex. *Cenp-HIK-TW complexes.* To test which regions of Cenp-H, Cenp-I and Cenp-K interact with each other and with Cenp-TW, the following fragments of Cenp-H, Cenp-I and Cenp-K were constructed: Cenp-I (residues 1–308) (Cenp-I$^N$), Cenp-H (residues 137–182) (Cenp-H$^C$), Cenp-H (residues 130–239) (Cenp-K$^C$) and combinations of Cenp-H, Cenp-I and Cenp-K, together with Cenp-TW were assembled into the pU1 plasmid for Multibac expression[34] for co-expression using the insect cell–baculovirus system. A double StrepII tag was added to the C terminus of Cenp-I.

To test the role of the positively charged DNA-binding groove of Cenp-N for Cenp-A nucleosome interactions, a total of 13 Arg and Lys mutations were introduced into *CHL4* (Cenp-N$^{Mut}$) by total gene synthesis (GeneArt/Thermo Fisher): chl4$^{K22S/K26S/R67S/K100S/ K103S/K105S/R198S/ K217S/K245S/K249S/K384S/ K401S/K403S}$. Cenp-N$^{Mut}$ was combined with Cenp-L to generate a Cenp-N$^{Mut}$-Cenp-L co-expression baculovirus.

The baculoviruses for expression of Cenp-OPQU+, Cenp-HIK–TW, Cenp-C and Cenp-LN were prepared for expression using the insect cell-baculovirus system[34].

The cDNA encoding *S. cerevisiae CSE4* (*S. cerevisiae CENP-A*), *H2A*, *H2B* and *H4* histone genes were synthesized (GeneArts/Thermo Fisher) with optimized codons for expression in *Escherichia coli* and were subsequently cloned into pET28A with a TEV protease cleavable N-terminal His$_6$ tag. For the recombinant Cse4 octamer (Cenp-A octamer), four expression cassettes for *CSE4*, *H2A*, *H2B* and *H4* histone genes were subsequently cloned into a single pET28 plasmid by USER methodology[34] for *E. coli* expression. For *S. cerevisiae* H3 octamer purification, *CSE4* was replaced by the *H3* gene. The Cenp-A L1 loop mutant (Cenp-A$^{L1}$, cse4$^{K172S/D173A/Q174A/D175S}$) and cse4$^{130–229}$ ($^{ΔN}$Cenp-A) were expressed to produce Cenp-A$^{L1}$ and $^{ΔN}$Cenp-A octamers and nucleosomes, respectively. The chimeric H3$^N$-Cenp-A histone comprises a fusion of residues 1–50 of H3 with residues 141–229 of *CSE4*. The H3$^N$-Cenp-A histone (molecular mass 15.74 kDa) was used to generate H3$^N$-Cenp-A$^{Nuc}$–Widom 601 by the same procedure as for Cenp-A$^{Nuc}$.
*Expression and purification.* Complexes of Cenp-OPQU+, Cenp-HIK, Cenp-HIK–TW, Cenp-LN and Cenp-C were expressed individually in High-5 insect cells (*Trichoplusia ni*: expression system). The High-5 insect cell line was not tested for mycoplasma contamination and was not authenticated. The cells were collected 48 h after infection. The lysate was loaded onto a Strep-Tactin column (Qiagen) and the complexes were eluted with 2.5 mM desthiobiotin (Sigma) in a buffer of 50 mM Tris.HCl (pH 8.0), 200 mM NaCl, 1 mM DTT. The StrepII-tag was cleaved using TEV protease overnight at 4 °C. The proteins and complexes were further purified on Resource Q anion-exchange and SEC in a buffer of 20 mM Hepes (pH 8.0), 200 mM NaCl, 2 mM DTT. Free Cenp-HIK was crosslinked using 0.05% glutaraldehyde for 8 min on ice and quenched with 50 mM Tris.HCl (pH 8.0), then further purified using Superose 6 SEC. The proteins and complexes were collected, concentrated, frozen in liquid nitrogen and stored at −80 °C. The stable 14-subunit CCAN complex was reconstituted by combining individually purified CCAN subcomplexes; Cenp-LN, Cenp-OPQU together with the budding yeast-specific Nkp1 and Nkp2 subunits (Cenp-OPQU+), Cenp-HIK–TW and Cenp-C.

For Cenp-HIK–TW assembly assays, a combination of full-length and either their N or C-terminal fragments of Cenp-I, Cenp-H and Cenp-K were co-expressed together with Cenp-T and Cenp-W or with Cenp-T$^{HFD}$ (residues 268–361) and Cenp-W. Affinity-purified complexes were analysed using SDS–PAGE analysis.

The *S. cerevisiae* Cenp-A octamer was prepared by co-expression of *CSE4*, *H2A*, *H2B* and *H4* in B834$^{Rare2}$ *E. coli* cells. The collected cell pellet was lysed in a buffer of 50 mM Tris.HCl (pH 8.0), 2 M NaCl. The Cenp-A octamer was isolated by Ni-NTA affinity chromatography, eluted with imidazole in 2 M NaCl buffer. The octamer was further purified by S200 SEC, concentrated to 3 mg ml$^{-1}$ in a buffer of 10 mM Tris.HCl (pH 7.5), 2 M NaCl, 1 mM EDTA and 2 mM DTT and frozen in liquid nitrogen and stored at −80 °C.

For DNA-fragment preparation, NEB Stable *E. coli* cells containing a plasmid with a multiple copy (20×) of the 147-bp Widom 601 sequence flanked by EcoRV sites in a pUC18 backbone (gift from F. Martino, MRC–LMB) were cultured in LB broth with ampicillin. The plasmid was isolated by using the Plasmid Giga Kit (Qiagen). The Widom 601 fragment was purified with a 1 ml resource Q anion-exchange chromatography column (GE Healthcare Life Sciences) after overnight digestion with EcoRV-HF (NEB). The purified DNA was precipitated, dissolved, buffer-exchanged and stored in a buffer of 2 M NaCl, 10 mM Tris.HCl (pH 7.5), 1 mM EDTA, 2 mM DTT at −20 °C. The *CEN3* DNA fragment was prepared by the primer-extension method. The two oligonucleotides used were: CEN3F, ATAAGTCACATGATGATATTTGATTT TATTATATTTTTAAAAAAAGTAAAAAATAA AAAGTAGTTTATTTTTAAA AAATAAAATTTAAAA and CEN3R, TTCAATGAAATATATATTTCTTA CTATTTCTTTTTTAACTTTCGGAAATCAAATACACTAATATTTTAAATTT TATTTTTTAAAAAATAAACTA (Sigma-Aldrich). The fragment was produced in a one-step extension at 68 °C for 1 min. The final product of the 153-bp *CEN3* (ATAAGTCACATGATGATATTT GATTTTATTATATTTTTAAAAAAAGT AAAAAATAAAAAGTAGTTTATTTTTAAAAAATAAAATTTAAAATATTAG TGTATTTGATTTCCGAAAGTTAAAAAAGAAATAGTAAGAAATATATATTT CATTGAA) fragment was purified using a 1-ml resource Q anion-exchange chromatography column and stored in a buffer of 2 M NaCl, 10 mM Tris.HCl (pH 7.5), 1 mM EDTA and 2 mM DTT at −20 °C.
*Cenp-A nucleosome and derivatives preparation.* Cenp-A, Cenp-A-L1$^{Mut}$, $^{ΔN}$Cenp-A, H3$^N$-Cenp-A and H3 histone octamers were wrapped by gradient dialysis from 2 M NaCl to 100 mM NaCl buffer with 10 mM Tris.HCl (pH 7.5), 1 mM EDTA and 2 mM DTT. Cenp-A octamer was mixed with either Widom 601 DNA or *CEN3* DNA at 7.8 μM concentration. The mixture in the dialysis tube was inserted into a 500-ml beaker containing 500 ml buffer of 2 M NaCl, 10 mM Tris.HCl (pH 7.5), 1 mM EDTA and 2 mM DTT. The NaCl concentration in the dialysis buffer was gradually decreased to 100 mM using an Akta pump at 1.5 ml min$^{-1}$ for 16 h at 4 °C. The mixture was further dialysed against the buffer of 100 mM NaCl, 10 mM Tris.HCl (pH 7.5), 1 mM EDTA, 2 mM DTT for 4 h at 4 °C. The Cenp-A nucleosome and derivatives were stored at 4 °C.
*Reconstitution of CCAN−Cenp-A nucleosome complex.* The CCAN−Cenp-A nucleosome complex was reconstituted by mixing purified Cenp-C and Cenp-LN with Cenp-A nucleosome followed by Cenp-HIK−TW and Cenp-OPQU+. The stoichiometry of CCAN subcomplexes to Cenp-A$^{Nuc}$ was adjusted so that CCAN subcomplexes were in excess, as judged by their separation from CCAN−Cenp-A$^{Nuc}$ by SEC. The mixed sample was dialysed overnight in a buffer of 10 mM Hepes (pH 8.0), 80 mM NaCl, 1 mM EDTA and 0.5 mM TCEP at 4 °C. CCAN−Cenp-A$^{Nuc}$ was purified by Superose 6 SEC. For cryo-EM analysis, CCAN−Cenp-A$^{Nuc}$ was crosslinked with 5 mM BS3 (Thermo Fisher Scientific) for 1 h on ice and quenched with 50 mM Tris and then subjected to further SEC with an Agilent Bio SEC-5 column (Agilent Technologies) before preparing cryo-EM grids. Mild crosslinking of CCAN–Cenp-A$^{Nuc}$ reduced dissociation of CCAN from Cenp-A$^{Nuc}$ during preparation of cryo-EM grids. To assess whether crosslinked created artefacts, we also collected a cryo-EM dataset using uncrosslinked CCAN–Cenp-A$^{Nuc}$.
**SEC analysis of CCAN−Cenp-A$^{Nuc}$ complexes.** To analyse the formation and stability of CCAN−Cenp-A$^{Nuc}$ complexes and mutants in CCAN and Cenp-A, all CCAN−Cenp-A$^{Nuc}$ complexes were assembled as above (with or without Cenp-C) and then applied to an Agilent Bio SEC-5 SEC column. The eluted fractions were analysed on SDS–PAGE gels and stained with Coomassie blue and ethidium bromide to detect proteins and DNA. For assembly of the CCAN−Cenp-A$^{Nuc}$ complexes, the concentration of Cenp-A$^{Nuc}$ was 1.6 μM, and that for the individual CCAN subcomplexes was 1.6 μM.
**SEC–MALS.** SEC–MALS was performed using a Wyatt MALS system. CCAN alone, uncrosslinked and BS3 crosslinked CCAN−Cenp-A$^{Nuc}$ complexes were injected onto an Agilent Bio SEC-5 column gel filtration column pre-equilibrated in 10 mM Hepes (pH 7.5), 80 mM NaCl, 1 mM EDTA and 0.5 mM TCEP. The light scattering and protein concentration at each point across the peaks in the chromatograph were used to determine the absolute molecular mass from the intercept of the Debye plot using Zimm's model as implemented in the ASTRA v.5.3.4.20 software (Wyatt Technologies). To determine inter-detector delay volumes, band-broadening constants and detector intensity normalization constants for the instrument, we used aldolase as a standard prior-to sample measurement. Data were plotted with the program Prism v.8.2.0 (GraphPad Software).
**Analytical ultracentrifugation.** Uncrosslinked and BS3 crosslinked CCAN−Cenp-A$^{Nuc}$ complex at approximately 1 mg ml$^{-1}$ in 10 mM Hepes (pH 7.5), 80 mM NaCl, 1 mM EDTA and 0.5 mM TCEP were subjected to velocity sedimentation at 40,000 r.p.m. at 4 °C in an An50Ti rotor using an Optima XL-I analytical ultracentrifuge (Beckman). The data were analysed in SEDFIT 16.1[35] using a *c(s)* distribution model. The partial-specific volumes (v-bar) were calculated using Sednterp (v.20130813 beta) (T. Laue, University of New Hampshire). The density and viscosity of the buffer were determined with a DMA 4500M density meter

(Anton Paar) and an AMVn viscometer (Anton Paar). Data were plotted with the program GUSSI[36].

**Micrococcal nuclease digestion assay.** Nucleosomes were digested for 40 min with 1 unit of MNase (NEB) per microgram of DNA at room temperature (22 °C). Reactions were terminated with the addition of excess EGTA. The digested nucleosome mixtures were loaded onto an agarose gel and stained to visualize the DNA.

**Yeast strains and growth analysis.** The *S. cerevisiae* strain with a *chl4* deletion and *cse4-R37A* mutation (*chl4Δ cse4-R37A*), AEY4992 (*MATα ade2-101 lys2 his3-11,15 trp1-1 leu2-3,112 ura3-1 can1-100 chl4Δ::kanMX cse4-R37A*) and wild-type *S. cerevisiae* strain (W303) (*MATα ade2-101 his3-11,15 trp1-1 leu2-3,112 ura3-1*) have previously been described and authenticated[27,37]. Yeast strains do not have mycoplasma and were not tested for mycoplasma contamination. Cenp-N[WT] and Cenp-N[Mut] strains were created by transforming AEY4992[27,37] with a 2μ origin plasmid pYes2 incorporating either *CHL4[WT]* or *chl4[Mut]* (*chl4[K22S/K26S/R67S/K100S/K103S/K105S/ R198S/K217S/K245S/K249S/K384S/K401S/K403S]*) with the native promoter of *CHL4*, a C-terminal double StrepII-tag on Chl4, and the *URA3* selection marker. The transformed cells were selected on synthetic medium lacking uracil, and the presence of the plasmid-encoded *CHL4* was verified by PCR using a primer pair over-spanning the *CHL4* and *URA3* genes. Cells were grown in drop-out uracil (SC-U) medium at 30 °C and spotted in tenfold dilution steps on YPED plates. The plates were incubated at either 30 °C or 37 °C for three days.

**Immunoprecipitation and western blotting for detecting Cenp-N expression in the *chl4Δ cse4-R37A* yeast.** Six litres of synthetic SC-U culture were inoculated with the *chl4Δ cse4-R37A* yeast strain transformed with the pYes2 plasmid expressing either wild type or mutant Cenp-N with a C-terminal double StrepII-tag (and empty vector control) and collected at $OD_{600\,nm}$ of approximately 0.8. Pelleted cells were lysed in buffer (50 mM Tris, pH 8.0, 300 mM NaCl, 1 mM EDTA and 1 mM DTT) and the cleared lysate was loaded onto a 1-ml Streptactin column. Fractions were eluted with 5 mM desthiobiotin and analysed by SDS–PAGE. Western blotting was performed with a Strep-tag antibody (MCA2489P, Bio-Rad) that detected the C-terminal double StrepII-tag on Cenp-N. Total protein was analysed by Coomassie blue staining for loading controls (normalized loading).

**Electron microscopy data collection.** Three microlitres of the CCAN−Cenp-A[Nuc] complex at a concentration of about 1 mg ml[−1] was applied to glow-discharged copper 300 mesh Quantifoil R1.2/1.3 holey carbon grids (Quantifoil Micro Tools) (no carbon support). The grids were flash-frozen by being plunged into liquid ethane using an FEI Vitrobot Mark IV (waiting time, 20 s, blotting time, 2 s). Cryo-EM image stacks were collected with Falcon III cameras in counting mode on four different FEI Titan Krios electron microscopes at a nominal magnification of 75 K (yielding pixel sizes of 1.065Å, 1.070 Å, 1.085 Å and 1.090 Å, respectively). The images were recorded at a dose rate of 0.6 electrons per pixel per second and the total exposure time was 60 s (75 frames) with the FEI automated low-dose data-collection program EPU. Defocus varied from −2.0 to −2.8 μm with an interval of 0.2 μm.

For the isolated Cenp-HIK sample, freshly purified Cenp-HIK complex was first visualized by negative-staining cryo-EM to check the sample quality. Aliquots of 3 μl samples at about 0.2 mg m[−1] were applied onto glow-discharged Quantifoil R1.2/1.3 300 mesh holey carbon grids. The grids were incubated for 30 s at 4 °C and 100% humidity and then blotted for 8 s and plunged into liquid ethane using an FEI Vitrobot III. Grids made in this way showed strong preferred orientation. To overcome this problem, we treated the Cenp-HIK complex with 0.025% glutaraldehyde for 10 min on ice before SEC purification. More views were observed after this treatment, allowing us to reconstruct the 3D structure.

For the isolated Cenp-HIK subcomplex, images were collected using EPU with a Falcon III detector in counting mode. Nine hundred and ten micrographs were collected using a dose rate of 0.5 electrons per pixel per second and a total exposure time of 60 s. Each micrograph was recorded into a movie stack of 75 frames. Calibrated physical pixel size is 1.38 Å per pixel.

**Image processing.** Movie frames were first aligned using MotionCor2[38]. CTF parameters were estimated with Gctf[39]. The initial template-free particle picking was performed with Gautomatch (developed by K. Zhang, https://www.mrc-lmb.cam.ac.uk/kzhang/Gautomatch/). Subsequent image processing was carried out using RELION 2.1 and RELION 3.0[40,41]. A subset of 556 micrographs (of 1,582) was used for Gautomatch template-free particle picking, and the resulting 119,143 coordinates were imported into RELION 2.1 for particle extraction and reference-free 2D classification. Selected averages from the 2D classification were used for an initial model reconstruction with SIMPLE-PRIME[42]. These 2D class averages were used for template-based particle auto-picking in Gautomatch for the entire dataset. The extracted particles were subject to 2 rounds of reference-free 2D classifications, resulting in a dataset of 1,385,496 particles from the combined total of 9,002 micrographs. A tandem cascade of 3D classifications against the model built with SIMPLE-PRIME[42] was performed, and initial iterations were performed without angular search restriction for each round of classification. After removing the bad particles, 424,577 particles were assigned to CCAN, whereas 193,882 were

assigned to the CCAN−Cenp-A[Nuc], which were used for the subsequent Bayesian polishing, multi-body refinement (MBR), and the final map refinement and atomic coordinate refinement. Beam-tilt parameters of the particles were estimated based on the individual dataset, and they were applied during the Bayesian polishing of each dataset in RELION 3.0. Refinements in 3D and MBRs were performed with the polished particle stacks after merging all the datasets. The dataset including all the particles generated the highest resolution reconstruction with an overall CCAN mask. The final resolutions for CCAN and CCAN−Cenp-A[Nuc] are 3.55 Å and 4.15 Å, respectively, based on the gold-standard Fourier shell correlation (FSC) = 0.143 criterion[43] (Extended Data Fig. 2d).

To identify (CCAN)$_2$–Cenp-A[Nuc] particles, five 2D classes, with 2D averages of CCAN-Cenp-A[Nuc] (Extended Data Fig. 2c) that showed smeared density in close proximity to Cenp-A[Nuc], were selected for further analyses. The selected particles (10, 553 particles) were subject to a tandem cassette of 2D classifications, resulting in 556 particles, which showed clear C2-symmetry 2D averages. These particles were re-extracted from the micrographs with a box size of 400 pixels to accommodate the bigger symmetric particles. The re-extracted particles were then subject to further 2D classification, and classified into 20 classes, generating the representative symmetric 2D averages shown in the red box of Extended Data Fig. 2c. The reprojections of the modelled (CCAN)$_2$–Cenp-A[Nuc] map (filtered to 20 Å resolution) were generated with relion_project. The projections are shown in Extended Data Fig. 10h. The small number of particles and highly preferred orientation on the cryo-EM grid (in the plane of the two-fold symmetry axis) precluded a 3D reconstruction.

**MBR.** To improve map resolution we performed MBR in RELION 3.0[41]. Two masks were generated. Mask1 comprised Cenp-LN-OPQU+, excluding Cenp-HIK. Mask2 comprised Cenp-HIK and portions of Cenp-N, L, O and P, (Extended Data Figs. 2h, i; 3b). The resultant maps were determined at 3.45 Å and 3.83 Å resolution, respectively. To further improve regions at the periphery of Cenp-OPQU+, partial signal subtracted particles (Cenp-HIK subtracted) were used for a second round of MBR. Mask3 included part of Cenp-N and N-terminal regions of Cenp-Q, Cenp-U, Nkp1 and Nkp2 with small regions of Cenp-O and Cenp-P. Mask4 comprised Cenp-OP, Cenp-LN and C-terminal regions of Cenp-QU, Nkp1 and Nkp2. MBR based on mask3 and mask4 resulted in 3.92 Å and 3.49 Å maps, respectively. The resultant maps derived using multi-body refinement based on the four masks showed substantially improved definition of cryo-EM densities and were used for model building (Extended Data Figs. 2d, h, i; 3b). Careful choice of the boundaries of mask2 was critical to optimizing the cryo-EM density quality for Cenp-HIK. Including specific regions of Cenp-N, L, O and P within mask2 was critical to generating maps that allowed side-chain definition of the coiled-coil regions of Cenp-H and Cenp-K (Extended Data Fig. 4a). This defined the correct assignment and polarity of these chains. MBR also improved definition of side chains in the base of Cenp-HIK. The subsequent MBR using mask3 and mask4 improved side chain definition for the peripheral regions of Cenp-OPQU+. Portions of the cryo-EM density map are shown in Extended Data Fig. 4. A 3D class (4% of total apo-CCAN) corresponding to dimeric apo-CCAN was determined at 9 Å resolution (Extended Data Fig. 3a).

For the uncrosslinked dataset, the same procedures were applied. A total of 123,215 particles from 1,586 micrographs were used for the final reconstruction of a map at 7.8 Å resolution for the CCAN−Cenp-A[Nuc] complex (Extended Data Fig. 5a).

For the isolated Cenp-HIK complex, the same procedure was applied. A total of 374,158 particles were used for the final reconstruction of a map at 4.3 Å resolution for Cenp-HIK complex.

Before visualization, a negative B factor determined with RELION 2.1 was applied to the density map for sharpening. The modulation transfer function of the detector was corrected in the post-processing step with RELION 3.0[40]. The local resolution was estimated with RELION 3.0[40].

**Model building and structure refinement.** *Apo-CCAN.* Cryo-EM density maps were visualized in COOT[44] and Chimera[45]. The crystal structure of *K. lactis* Cenp-OPQ (PDB: 5MU3)[46] (equivalent to *S. cerevisiae* Cenp-O residues 159–362, *S. cerevisiae* Cenp-P residues 148–361 and *S. cerevisiae* Cenp-Q residues 320–342) and structures of *S. cerevisiae* Cenp-N (residues 374–450), Cenp-L (PDB: 4JE3)[47] and human Cenp-N N-terminal domain (NTD) (PDB: 6EQT)[29] (equivalent to residues 12–260 of *S. cerevisiae* Cenp-N) were fitted into the cryo-EM density maps of apo-CCAN, with refitting and mutating to the *S. cerevisiae* sequence for Cenp-N[NTD], Cenp-O, Cenp-P and Cenp-Q. On the basis of the good quality of the cryo-EM densities, atomic models of Nkp1, Nkp2, Cenp-U, Cenp-Q, Cenp-H (residues 7–136), Cenp-I (residues 321–728) and Cenp-K (residues 4–128) and the interdomain region of Cenp-N (residues 261 to 373) were built de novo. Only short stretches of Cenp-Q (residues 161–216) and Cenp-U (residues 131–155) were built as polyAla (Extended Data Table 2). The secondary-structural and disordered regions of the protein sequences were analysed with PHYRE2[48] and PSIPred[49]. A model for the Cenp-HIK head domain was based on the crystal structure of

regions of the Cenp-HIK assembly from *Chaetomium thermophilum* and *Thielavia terrestris* (PDB: 5Z08)[16] corresponding to *S. cerevisiae* Cenp-H (residues Asp143 to Ile181), Cenp-I (residues Leu5 to Ala241) and Cenp-K (residues Ala136 to Thr236) and derived using PHYRE2[48]. The 3.5 Å monomeric free CCAN coordinates were rigid-body-docked into the cryo-EM map The Cenp-HIK head domain was fitted to cryo-EM density of the dimeric apo-CCAN. A linker region that connects Cenp-N[NTD] with Cenp-N[CTD], not present in crystal structures, was built de novo.

*CCAN−Cenp-A[Nuc]*. The CCAN-complex model was then fit into the CCAN−Cenp-A[Nuc] cryo-EM map. The nucleosome was modelled on the *S. cerevisiae* H3 nucleosome (PDB: 1ID3)[50] with *S. cerevisiae* Cenp-A modelled on human Cenp-A (PDB: 3AN2)[22] and mutated to the *S. cerevisiae* Cenp-A sequence, and the Widom 601 DNA sequence (PDB: 3LZ0)[51]. The Cenp-C model (PDB: 4X23)[30] in the centromeric nucleosome was rigid body-docked into the cryo-EM density.

The apo-CCAN and CCAN−Cenp-A[Nuc] models (excluding the Cenp-HIK head domains) were optimized by several rounds of real-space refinement using PHENIX (phenix.real_space_refine)[52]. Standard stereochemical and secondary structural constraints were applied during the real-space refinement. The final models were evaluated with COOT[44], PHENIX[52] and MolProbity (http://molprobity.biochem.duke.edu/)[53]. Figures were prepared using ChimeraX[54], Chimera[45], and PyMOL (Molecular Graphics System, 2.0.3, Schrodinger). Details of the fitted and refined coordinates in Extended Data Table 2. Multiple sequence alignments were performed and displayed using JALVIEW[55].

**XL-MS.** To assess the validity of our structure, we performed XL-MS analysis of the complexes[56]. Three independent crosslinking reactions were performed for each sample. The CCAN or CCAN–Cenp-A[Nuc] complexes in 20 mM Hepes pH 7.5, 80 mM NaCl and at a concentration of 3 mg ml$^{-1}$ were crosslinked with 1 mM DSSO for 15 min at room temperature. Each reaction was quenched with Tris.HCl (pH 8.0) to 50 mM and supplemented with urea to 8 M. The samples were reduced by addition of DTT at a final concentration of 10 mM for 1 h at room temperature, and alkylated for 0.5 h at room temperature in the dark by addition of iodoacetamide to 50 mM. Protein digestion was performed with Lys-C at an enzyme-to-protein ratio of 1:75 (w:w) at 30 °C for 3 h, then the samples were diluted in 50 mM ammonium bicarbonate and further digested with trypsin at an enzyme-to-protein ratio of 1:75 (w:w) at 37 °C for 16 h. The digested samples were acidified with formic acid to 1%, desalted using home-made C18 stage tips, dried and stored at −80 °C for further use.

Each sample was analysed by liquid chromatography with tandem mass spectrometry using an Agilent 1290 Infinity System (Agilent Technologies) in combination with an Orbitrap Fusion Lumos (Thermo Scientific). Reverse-phase chromatography was carried out using a 100-μm inner diameter, 2-cm trap column (packed in-house with ReproSil-Pur C18-AQ, 3 μm) coupled to a 75-μm inner diameter, 50-cm analytical column (packed in-house with Poroshell 120 EC-C18, 2.7 μm) (Agilent Technologies). Mobile-phase solvent A consisted of 0.1% formic acid in water, and mobile-phase solvent B consisted of 0.1% formic acid in 80% acetonitrile. A 180-min gradient was used, and start and end percentage buffer B adjusted to maximize the sample separation.

Mass spectrum acquisition was performed using the MS2_MS3 strategy: the MS1 scan was recorded in Orbitrap at a resolution of 60,000, the selected precursors were fragmented in MS2 with CID and the crosslinker signature peaks recorded at a resolution of 30,000. The fragments displaying the mass difference specific for DSSO were further fragmented in a MS3 scan in the ion trap[57]. Each sample was analysed with Proteome Discoverer 2.3 (v.2.3.0.522) with the XlinkX nodes integrated[57] and searching against databases generated after bottom-up analysis of the samples. The crosslink output (Supplementary Tables 1, 2) was subsequently visualized using the xVis[58] web tool and the crosslinks mapped onto the cryo-EM structures of CCAN and CCAN–Cenp-A using PyMOL (Molecular Graphics System, 2.0.3, Schrodinger) (Extended Data Fig. 6e–g). The XL-MS raw files, the associated output and databases were deposited with the ProteomeXchange Consortium[59].

**Modelling the CCAN–Cenp-A[Nuc]–CBF3–*Cen3* complex.** To model CCAN and CBF3 simultaneously bound to the Cenp-A nucleosome, we docked the free unwrapped DNA duplex of the CCAN–Cenp-A[Nuc] complex onto the CBF3–*Cen3* coordinates (PDB: 6GYS)[60], matching the minor and major grooves of both complexes. To avoid overlap of CBF3 and CCAN, the dyad symmetry axis of the Cenp-A nucleosome was positioned seven nucleotides upstream of the midpoint of CDEII of the *Cen3* sequence.

**Modelling human and *S. pombe* Cenp-LN complexes.** To generate the human Cenp-LN complex we used residues 1–207 from PDB 6EQT[29], and modelled residues 208–338 and Cenp-N by one-to-one threading in PHYRE2[48] using *S. cerevisiae* Cenp-LN as a template. *S. pombe* Cenp-LN was modelled with PHYRE2[48] using *S. cerevisiae* Cenp-LN as a template. The electrostatic potential of *S. cerevisiae*, *S. pombe* and *Homo sapiens* Cenp-LN complexes were calculated and displayed in PyMOL (Molecular Graphics System, 2.0.3, Schrodinger).

**Reporting Summary.** Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

## Data availability

Electron microscopy maps have been deposited with the Electron Microscopy Data Bank with accession codes EMD-4580 (CCAN), EMD-4579 (CCAN-Cenp-A[Nuc]), EMD-4581 (mask1) and EMD-4971 (mask2). Protein coordinates have been deposited with the PDB with accession codes 6QLE (CCAN), 6QLD (CCAN-Cenp-A[Nuc]) and 6QLF (mask1). The XL-MS raw files, the associated output and databases have been deposited through the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier PXD013769. Other data are available upon reasonable request.

34. Zhang, Z., Yang, J. & Barford, D. Recombinant expression and reconstitution of multiprotein complexes by the USER cloning method in the insect cell-baculovirus expression system. *Methods* **95**, 13–25 (2016).
35. Schuck, P. On the analysis of protein self-association by sedimentation velocity analytical ultracentrifugation. *Anal. Biochem.* **320**, 104–124 (2003).
36. Brautigam, C. A. Calculations and publication-quality illustrations for analytical ultracentrifugation data. *Methods Enzymol.* **562**, 109–133 (2015).
37. Samel, A., Cuomo, A., Bonaldi, T. & Ehrenhofer-Murray, A. E. Methylation of CenH3 arginine 37 regulates kinetochore integrity and chromosome segregation. *Proc. Natl Acad. Sci. USA* **109**, 9029–9034 (2012).
38. Zheng, S. Q. et al. MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat. Methods* **14**, 331–332 (2017).
39. Zhang, K. Gctf: real-time CTF determination and correction. *J. Struct. Biol.* **193**, 1–12 (2016).
40. Fernandez-Leiro, R. & Scheres, S. H. W. A pipeline approach to single-particle processing in RELION. *Acta Crystallogr. D* **73**, 496–502 (2017).
41. Nakane, T., Kimanius, D., Lindahl, E. & Scheres, S. H. Characterisation of molecular motions in cryo-EM single-particle data by multi-body refinement in RELION. *eLife* **7**, e36861 (2018).
42. Elmlund, H., Elmlund, D. & Bengio, S. PRIME: probabilistic initial 3D model generation for single-particle cryo-electron microscopy. *Structure* **21**, 1299–1306 (2013).
43. Chen, S. et al. High-resolution noise substitution to measure overfitting and validate resolution in 3D structure determination by single particle electron cryomicroscopy. *Ultramicroscopy* **135**, 24–35 (2013).
44. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D* **66**, 486–501 (2010).
45. Yang, Z. et al. UCSF Chimera, MODELLER, and IMP: an integrated modeling system. *J. Struct. Biol.* **179**, 269–278 (2012).
46. Schmitzberger, F. et al. Molecular basis for inner kinetochore configuration through RWD domain–peptide interactions. *EMBO J.* **36**, 3458–3482 (2017).
47. Hinshaw, S. M. & Harrison, S. C. An Iml3–Chl4 heterodimer links the core centromere to factors required for accurate chromosome segregation. *Cell Rep.* **5**, 29–36 (2013).
48. Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N. & Sternberg, M. J. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protoc.* **10**, 845–858 (2015).
49. Buchan, D. W., Minneci, F., Nugent, T. C., Bryson, K. & Jones, D. T. Scalable web services for the PSIPRED protein analysis workbench. *Nucleic Acids Res.* **41**, W349–W357 (2013).
50. Luger, K., Mäder, A. W., Richmond, R. K., Sargent, D. F. & Richmond, T. J. Crystal structure of the nucleosome core particle at 2.8 A resolution. *Nature* **389**, 251–260 (1997).
51. Vasudevan, D., Chua, E. Y. D. & Davey, C. A. Crystal structures of nucleosome core particles containing the '601' strong positioning sequence. *J. Mol. Biol.* **403**, 1–10 (2010).
52. Adams, P. D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D* **66**, 213–221 (2010).
53. Chen, V. B. et al. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D* **66**, 12–21 (2010).
54. Goddard, T. D. et al. UCSF ChimeraX: meeting modern challenges in visualization and analysis. *Protein Sci.* **27**, 14–25 (2018).
55. Waterhouse, A. M., Procter, J. B., Martin, D. M., Clamp, M. & Barton, G. J. Jalview version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics* **25**, 1189–1191 (2009).
56. Liu, F. & Heck, A. J. Interrogating the architecture of protein assemblies and protein interaction networks by cross-linking mass spectrometry. *Curr. Opin. Struct. Biol.* **35**, 100–108 (2015).
57. Liu, F., Lössl, P., Scheltema, R., Viner, R. & Heck, A. J. R. Optimized fragmentation schemes and data analysis strategies for proteome-wide cross-link identification. *Nat. Commun.* **8**, 15473 (2017).
58. Grimm, M., Zimniak, T., Kahraman, A. & Herzog, F. xVis: a web server for the schematic visualization and interpretation of crosslink-derived spatial restraints. *Nucleic Acids Res.* **43**, W362–W369 (2015).
59. Vizcaíno, J. A. et al. ProteomeXchange provides globally coordinated proteomics data submission and dissemination. *Nat. Biotechnol.* **32**, 223–226 (2014).
60. Yan, K., Zhang, Z., Yang, J., McLaughlin, S. H. & Barford, D. Architecture of the CBF3–centromere complex of the budding yeast kinetochore. *Nat. Struct. Mol. Biol.* **25**, 1103–1110 (2018).
61. Hinshaw, S. M., Dates, A. N. & Harrison, S. C. The structure of the yeast Ctf3 complex. *eLife* **8**, e48215 (2019).
62. Akiyoshi, B. et al. Tension directly stabilizes reconstituted kinetochore-microtubule attachments. *Nature* **468**, 576–579 (2010).
63. Gonen, S. et al. The structure of purified kinetochores reveals multiple microtubule-attachment sites. *Nat. Struct. Mol. Biol.* **19**, 925–929 (2012).

**Extended Data Fig. 1 |** See next page for caption.

**Extended Data Fig. 1 | Reconstituted *S. cerevisiae* CCAN–Cenp-A^Nuc complexes. a**, Size-exclusion chromatogram profiles (Agilent Bio SEC-5 column) for (i) CCAN, (ii) CCAN–Cenp-A nucleosome (with Widom 601) complex, (iii) Cenp-A nucleosome (with Widom 601), (iv) H3 nucleosome (with Widom 601) and (v) H3^N-Cenp-A^Nuc (with Widom 601). **b**, Comparative size-exclusion chromatogram profiles (Agilent Bio SEC-5 column) for CCAN–Cenp-A^Nuc with the Cenp-A nucleosome wrapped with either the (i) 147-bp Widom 601 positioning sequence (CCAN–Cenp-A^Nuc (Widom 601) as in **a**) or (ii) a 153-bp *S. cerevisiae* centromeric *Cen3* sequence (CCAN–Cenp-A^Nuc (*Cen3*)). Both complexes eluted at the same volume. CCAN and the H3 nucleosome do not form a complex (iii). **c**, Coomassie-blue-stained SDS–PAGE of the 14-subunit CCAN complex. **d**, Coomassie-blue-stained SDS–PAGE gel of Cenp-A^Nuc (Widom 601). Lane E32, ethidium bromide-stained gel of fraction 32. **e**, CCAN–Cenp-A^Nuc (Widom 601) complex. Lane E13, ethidium-bromide-stained gel of fraction 13. Size-exclusion chromatograms are shown in **a**. **f**, SDS–PAGE gel of CCAN and H3 nucleosome (Widom 601) SEC run shown in **b**. **g–j**, Coomassie-blue-stained SDS–PAGE gels of various Cenp-H, I and K segments co-expressed with Cenp-TW and purified with a double Strep tag on the tagged Cenp-I subunit (\*). **j**, The HFDs of Cenp-TW (Cenp-T^HFDW) interact with the Cenp-HIK^Head. These results confirm the assignments of the Cenp-H, K and I subunits in our cryo-EM maps. **k**, Schematic of the organization of CCAN–Cenp-A^Nuc subunits and subcomplexes and connections to the outer kinetochore Mis12 and Ndc80 complexes. Lines indicate subcomplex connections. The two pathways connecting Cenp-A^Nuc to the Ndc80 complex and microtubules are indicated as P1 and P2 (thick lines to Ndc80). Subunits of the essential P1 pathway are labelled black and indicated with blue shading, whereas subunits of the non-essential P2 pathway are labelled white and indicated with yellow shading. The P2 pathway becomes essential when the P1 pathway is defective through defects in Dsn1 phosphorylation[9]. The experiments shown in **a–j** were performed independently in triplicate with similar results. For gel source data, see Supplementary Fig. 1.
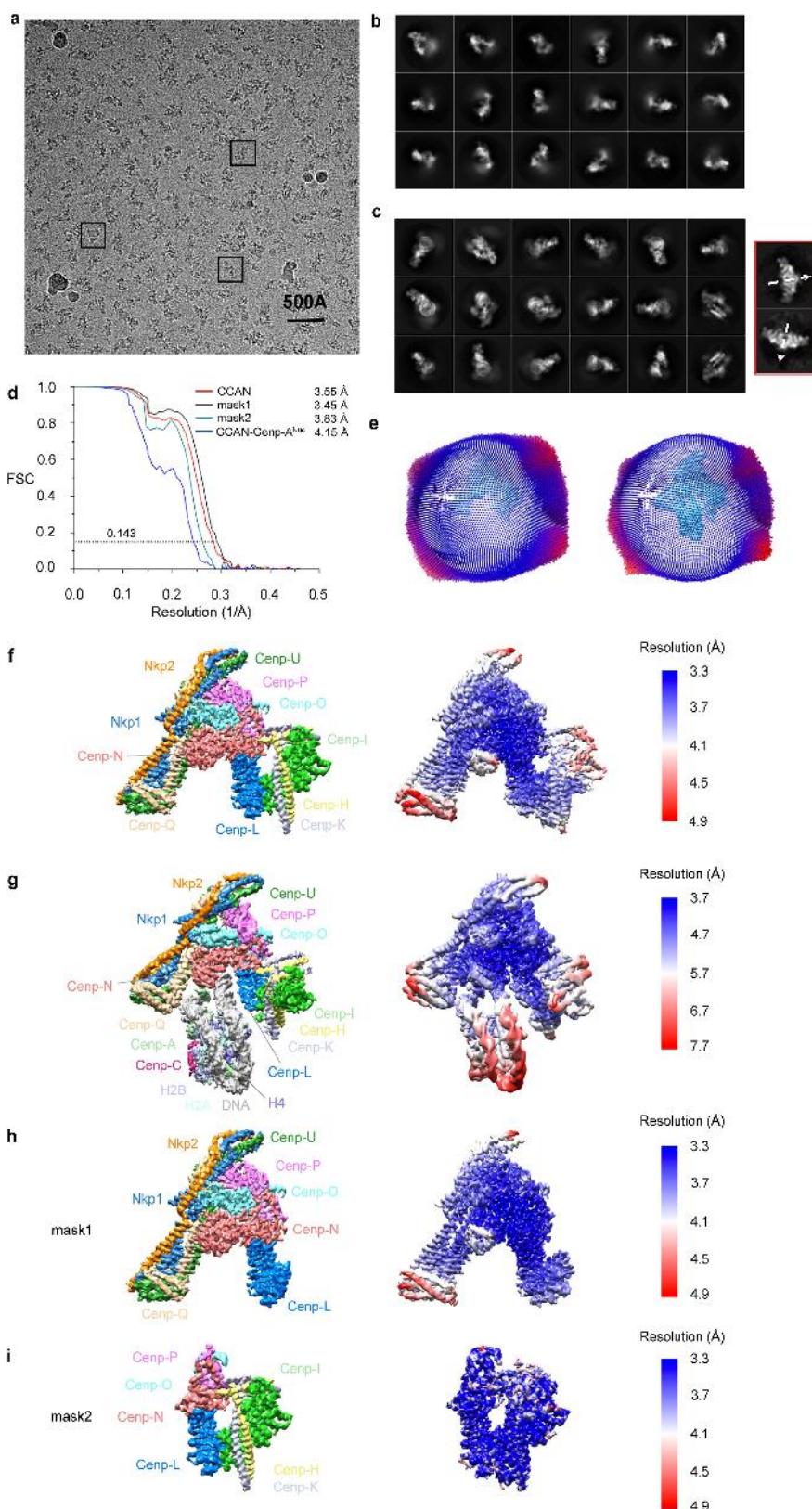
**Extended Data Fig. 2 |** See next page for caption.

**Extended Data Fig. 2 | Cryo-EM data of the *S. cerevisiae* CCAN–Cenp-A$^{Nuc}$ complex. a**, A typical cryo-electron micrograph of CCAN–Cenp-A$^{Nuc}$, representative of 9,002 micrographs. **b**, Galleries of 2D classes of CCAN, representative of 100 2D classes. **c**, Galleries of 2D classes of CCAN–Cenp-A$^{Nuc}$, representative of 150 2D classes. The 2D class averages for the C2-symmetric (CCAN)$_2$–Cenp-A$^{Nuc}$ complex viewed in the plane of the C2-symmetry axis are outlined in red. Only a few views were observed, precluding a 3D reconstruction. Cryo-EM grids partially destabilize CCAN–Cenp-A$^{Nuc}$ interactions, resulting in a very low abundance of (CCAN)$_2$–Cenp-A$^{Nuc}$ particles (about 0.03% of total). The two-fold symmetry axes of the (CCAN)$_2$–Cenp-A$^{Nuc}$ complex are shown as dashed arrows. Experiments for data in **b** and **c** were performed independently 12 times with similar results. **d**, FSC curves shown for the cryo-EM reconstructions of CCAN–Cenp-A$^{Nuc}$ complexes: apo-CCAN, mask1 (Cenp-OPQU+, Cenp-LN), mask2 (Cenp-HIK, Cenp-LN, sub-Cenp-OP), CCAN–Cenp-A$^{Nuc}$. Mask1 and mask2 used for MBR are defined in **h** and **i** and Methods. **e**, Angular distribution plot of CCAN–Cenp-A$^{Nuc}$ particles. **f**, Local resolution map of CCAN. **g**, Local resolution map of CCAN–Cenp-A$^{Nuc}$. **h**, Local resolution map of mask1 (Cenp-OPQU+, Cenp-LN). **i**, Local resolution map of mask2 (Cenp-HIK, Cenp-LN, sub-Cenp-OP).
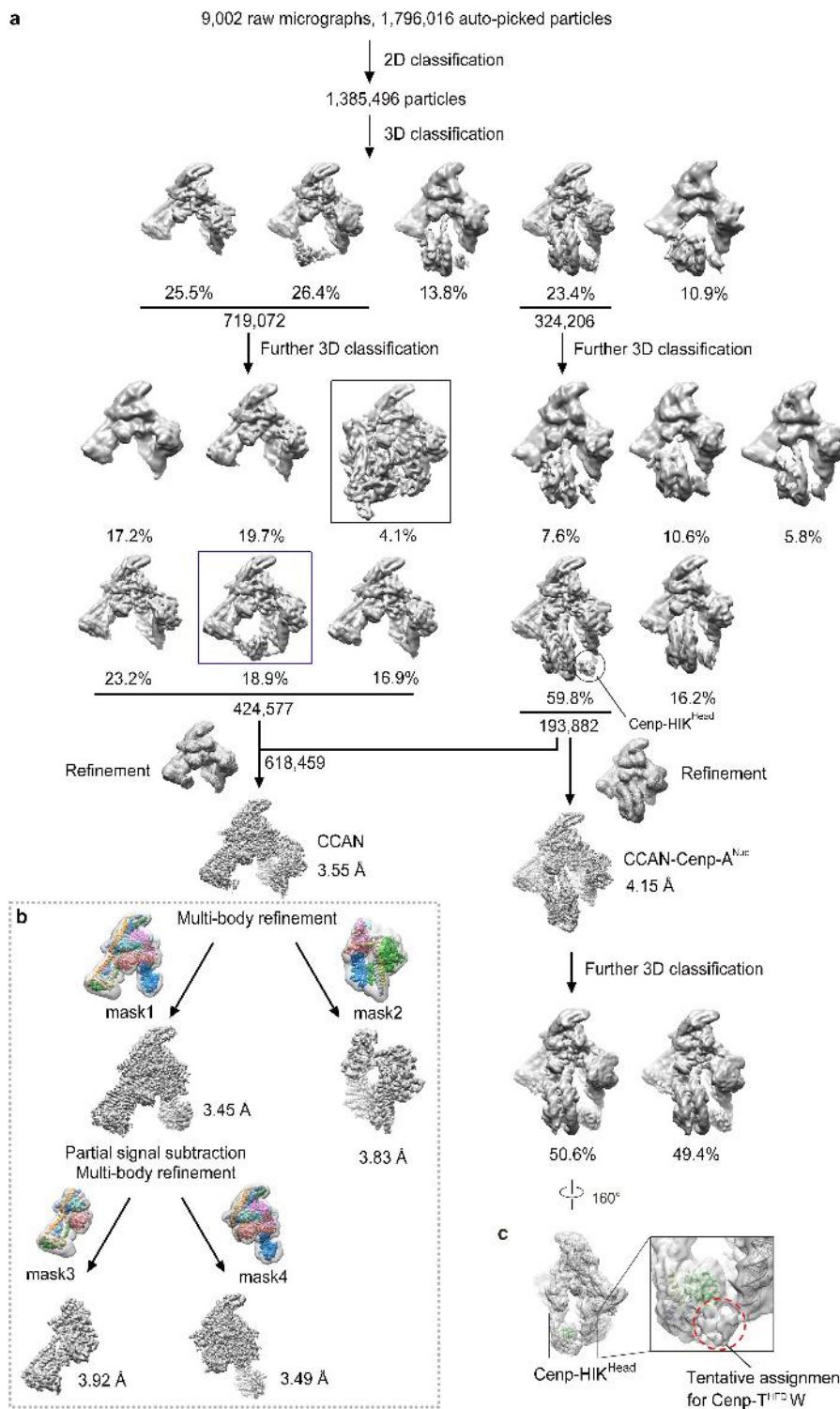
**Extended Data Fig. 3 | Workflow of 3D classification of the CCAN–Cenp-A$^{Nuc}$ cryo-EM dataset. a**, After initial 2D classification, about 1.4 million particles were sorted by 3D classification into apo-CCAN (52%) and the CCAN–Cenp-A$^{Nuc}$ complex (48%). For apo-CCAN, 4% existed as dimers (black box) and 19% showed an ordered head-group (Cenp-HIK$^{Head}$) for the Cenp-HIK–TW subcomplex (blue box). A mask was applied to the CCAN–Cenp-A$^{Nuc}$ cryo-EM map to exclude the structurally variable Cenp-HIK$^{Head}$ domain for reconstruction of the 4.15 Å structure. **b**, Details of the four masks used for MBR. **c**, A small 3D class of CCAN–Cenp-A$^{Nuc}$, revealing density attached to Cenp-HIK$^{Head}$ contacting the DNA gyre of Cenp-A$^{Nuc}$ was assigned as Cenp-T$^{HFD}$W.
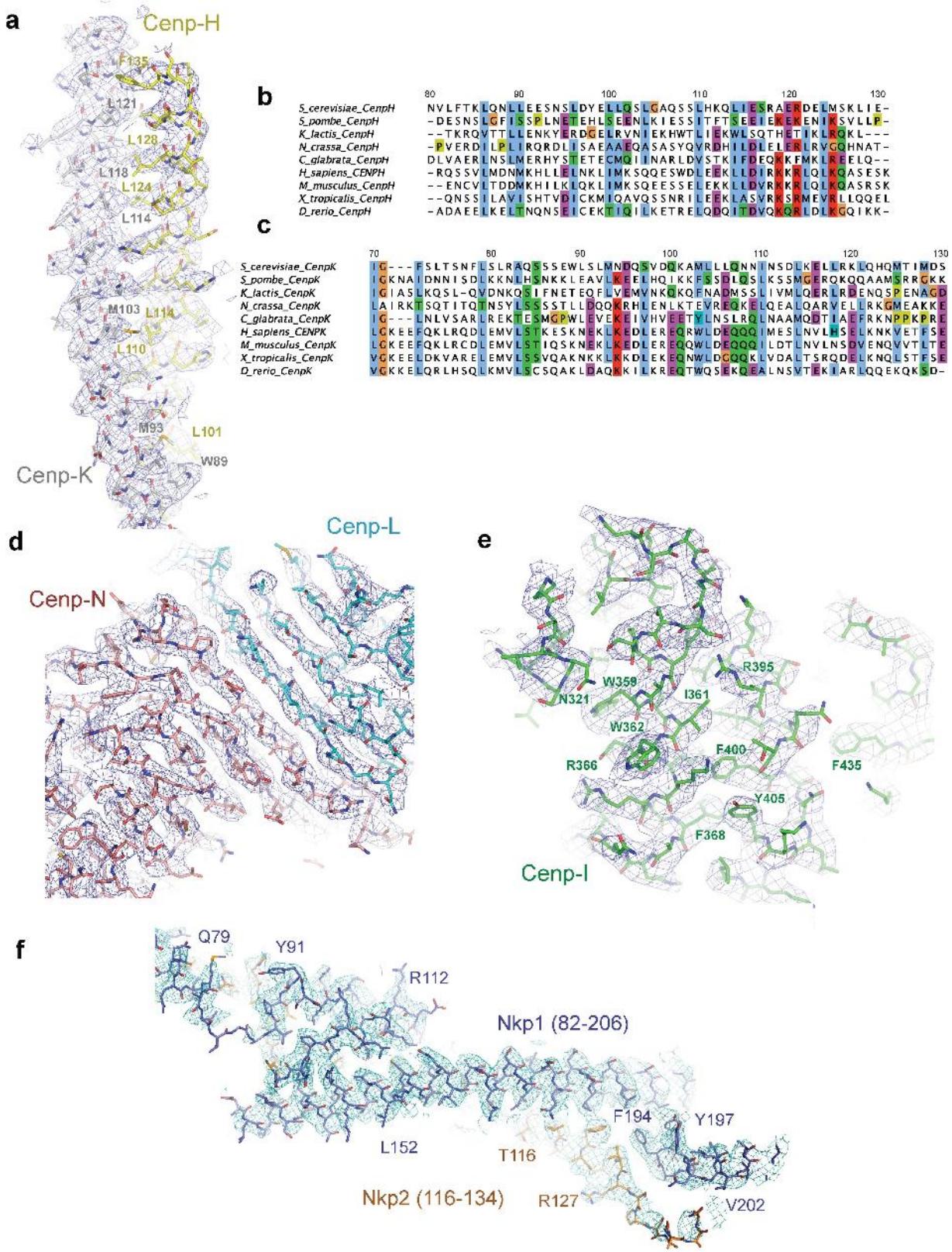
**Extended Data Fig. 4 | Cryo-EM density maps of apo-CCAN. a**, Portion of cryo-EM map for the coiled coils of Cenp-H and Cenp-K. A selection of highly conserved intersubunit residues defined in **b** and **c** are labelled. These residues are well defined in the cryo-EM density, consistent with the structure. **b**, **c**, Multiple sequence alignment of the coiled-coil regions of Cenp-H (**b**) and Cenp-K (**c**). **d**–**f**, Portions of cryo-EM maps for Cenp-LN (**d**), Cenp-I (**e**) and Nkp1–Nkp2 (**f**). The chain assignments and polarity of Cenp-H, Cenp-I and Cenp-K of our structure agree with the cryo-EM structure of yeast Ctf3 (PDB: 6OUA)[61].
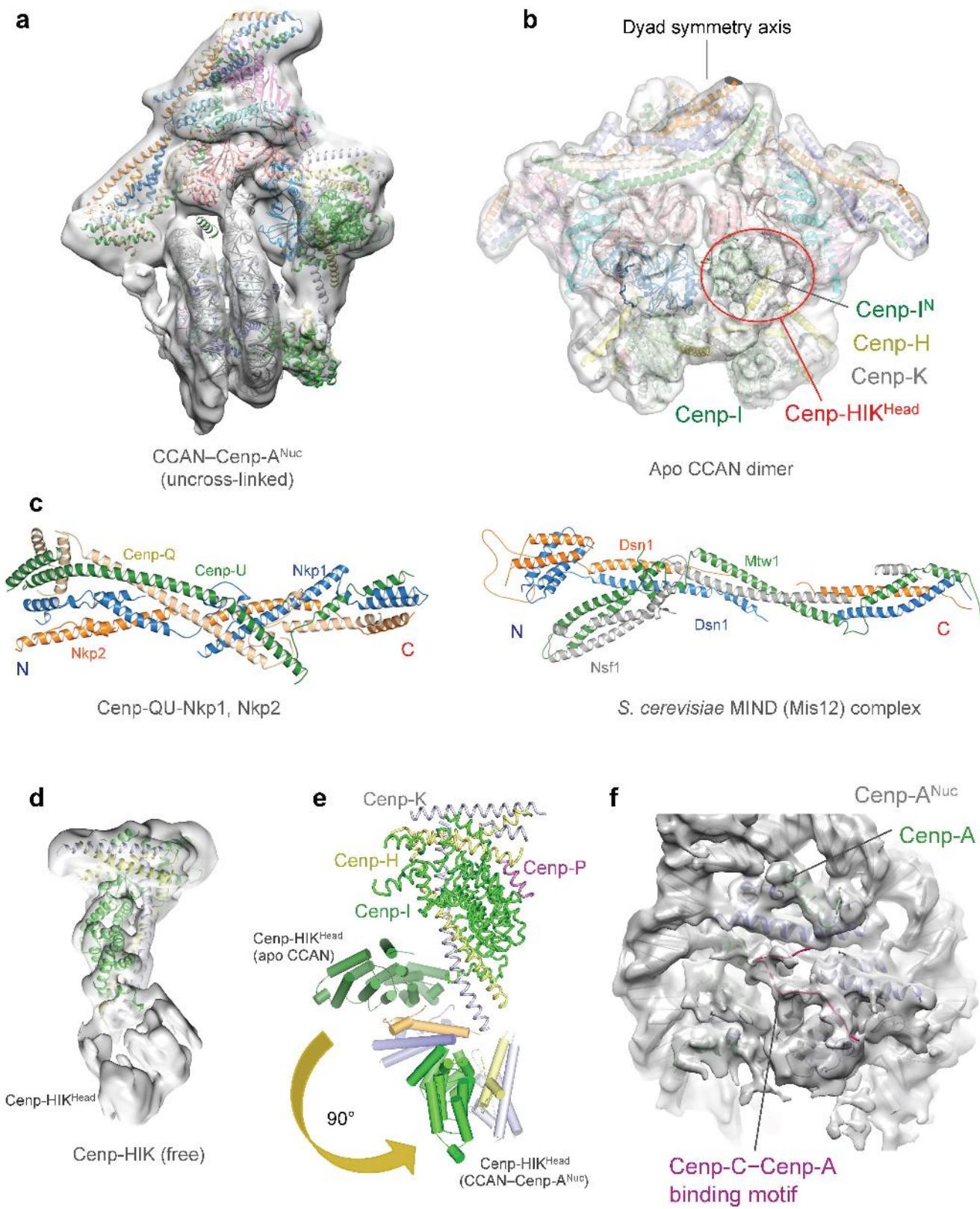
**Extended Data Fig. 5 | Cryo-EM densities of CCAN and CCAN–Cenp-A$^{Nuc}$ complexes. a**, Cryo-EM reconstruction of CCAN–Cenp-A$^{Nuc}$ from uncrosslinked sample at 8.6 Å resolution. **b**, Cryo-EM map of dimeric CCAN (also Extended Data Fig. 3a, black box). Subunits are colour-coded as in Fig. 1. The 3.5 Å monomeric free CCAN coordinates were rigid-body-docked into the cryo-EM map. **c**, Cartoon representation of the *S. cerevisiae* MIND complex[15] (right), showing a notable similarity to the coiled coils of Cenp-QU–Nkp1–Nkp2 of CENP-OPQU+ (left). **d**, View of the 4.7 Å resolution cryo-EM map of free Cenp-HIK with fitted coordinates from CCAN. **e**, In the context of CCAN, Cenp-HIK$^{Head}$ rotates to accommodate Cenp-A$^{Nuc}$. The two conformations of Cenp-HIK from the apo-CCAN and CCAN–Cenp-A$^{Nuc}$ complexes were superimposed onto their rigid portion of Cenp-HIK (C-terminal region of Cenp-I is shown for apo-CCAN) to indicate the conformational variability of Cenp-HIK$^{Head}$ between the two states. Subunits of Cenp-HIK$^{Head}$ of CCAN–Cenp-A$^{Nuc}$ are coloured lighter. **f**, Cryo-EM density of Cenp-A$^{Nuc}$ showing the Cenp-C motif of Cenp-C.
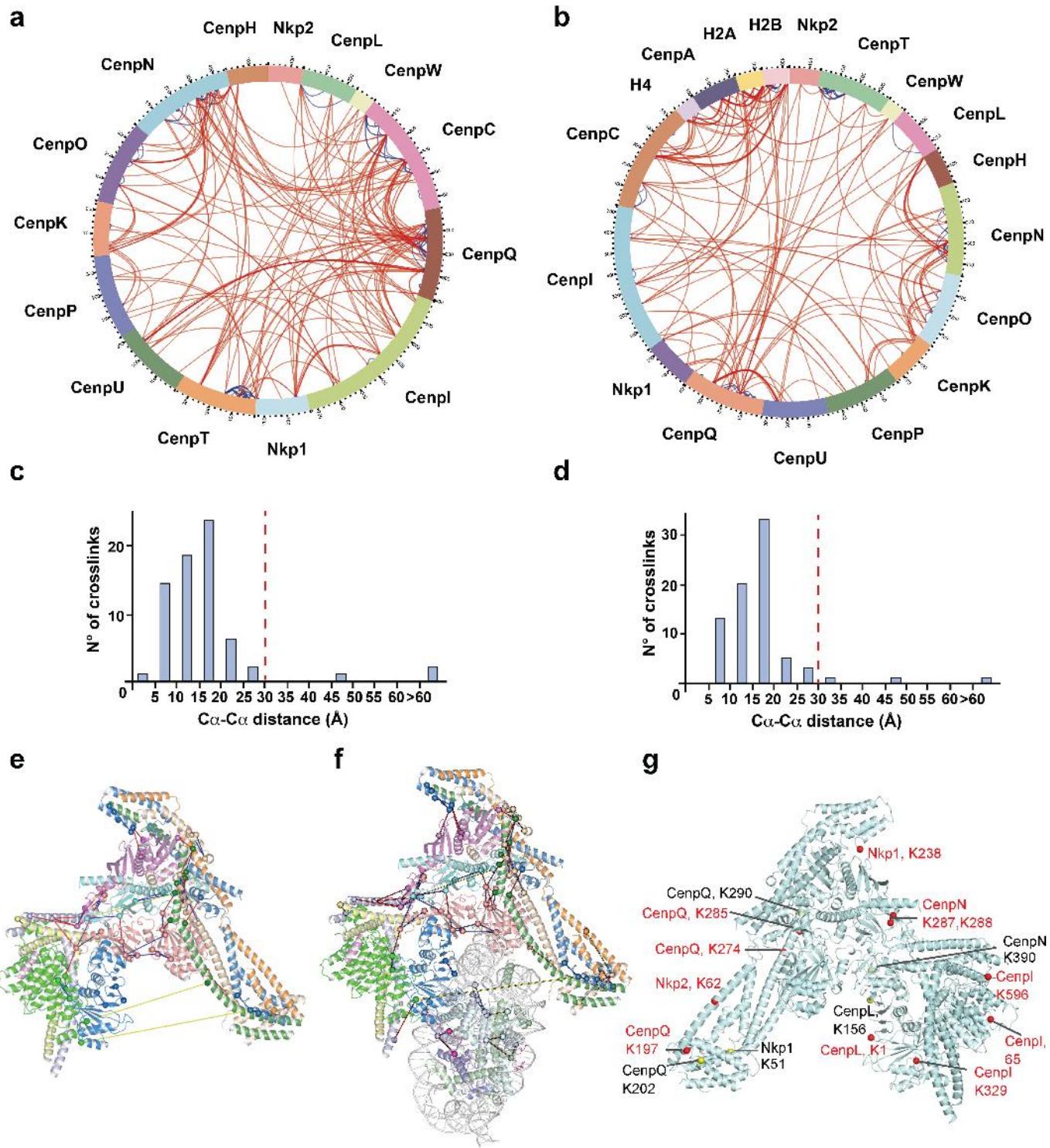
**Extended Data Fig. 6 | XL-MS analysis of the CCAN and CCAN–Cenp-A$^{Nuc}$ complexes. a**, **b**, Circular plots displaying all the identified crosslinks for CCAN (**a**) and CCAN–Cenp-A$^{Nuc}$ (**b**). Inter- and intra-subunit crosslinks are indicated in red and blue, respectively **c**, **d**, Histogram plots showing the $C_\alpha$–$C_\alpha$ distance distribution of the crosslinks that could be mapped onto the CCAN (**c**) and CCAN–Cenp-A$^{Nuc}$ structures (**d**). Ninety-five per cent of the mapped crosslinks satisfy the crosslinker-imposed distance restraint of 30 Å indicated with a dashed red line. **e**, **f**,

Crosslinks mapped onto the CCAN (**e**) and CCAN–Cenp-A$^{Nuc}$ complex (**f**). Inter and intra-subunit crosslinks are indicated in red and blue, respectively. Crosslinks exceeding the crosslinker-imposed distance restraint of 30 Å are indicated in yellow. **g**, Residues on CCAN shown by XL-MS that crosslink with Cenp-C are indicated on the CCAN structure. Red spheres, crosslinks in the CCAN–Cenp-A$^{Nuc}$ complex; yellow spheres, additional crosslinks unique to apo-CCAN. The experiments shown in **a** and **b** were performed independently in triplicate with similar results.
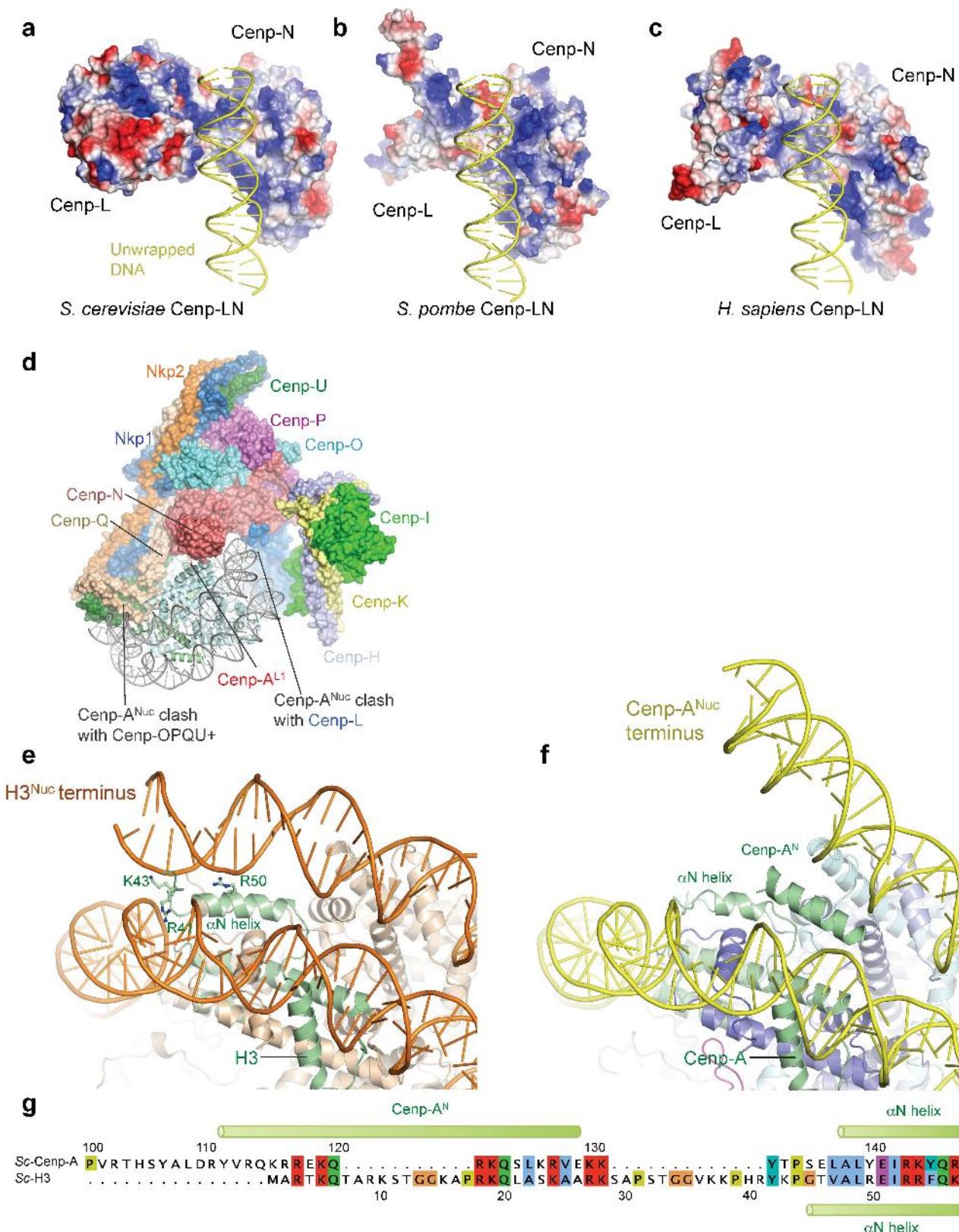
**Extended Data Fig. 7 | The *S. cerevisiae* Cenp-A$^{Nuc}$ nucleosome is unwrapped. a–c,** The positively charged electrostatic potential of the DNA-binding groove of Cenp-LN subcomplex is conserved in *S. cerevisiae*, *S. pombe* and *H. sapiens*. *S. pombe* and *H. sapiens* are represented by modelled structures. **d,** Cenp-N interacts with *S. cerevisiae* Cenp-A$^{Nuc}$ in the context of CCAN differently from the interaction of free human Cenp-N with Cenp-A nucleosome. The Cenp-N subunit of the human Cenp-N–Cenp-A nucleosome structure (PDB: 6C0W[29]) was superimposed onto Cenp-N of the *S. cerevisiae* CCAN–Cenp-A$^{Nuc}$ structure. In this mode of Cenp-N–Cenp-A$^{Nuc}$ interactions, Cenp-A$^{Nuc}$ would clash with Cenp-OPQU+ and Cenp-N of CCAN. **e, f,** Structures of *S. cerevisiae* H3$^{Nuc}$ (PDB: 1ID3[24]) (**e**) and Cenp-A$^{Nuc}$ (**f**, this work). **g,** Sequence alignment of the N-terminal regions of *S. cerevisiae* H3 and Cenp-A (Cse4) histones. For the chimeric H3$^N$–Cenp-A$^{Nuc}$, residues 1–50 of *S. cerevisiae* H3 were substituted for residues 1–140 of *S. cerevisiae* Cenp-A. A similar approach was used for vertebrate Cenp-A$^{Nuc}$ (ref. [23]).
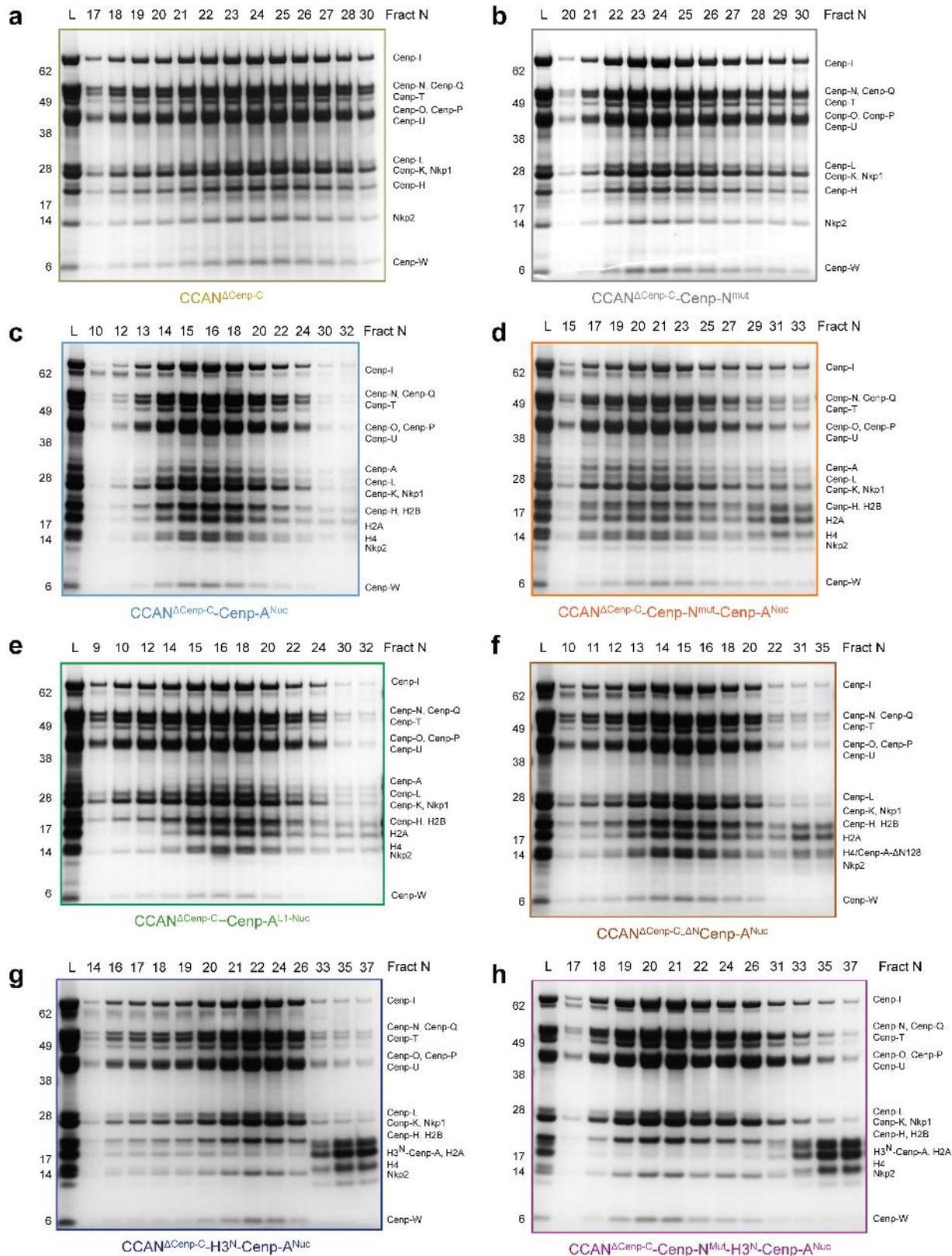
**Extended Data Fig. 8 | SDS–PAGE of CCAN$^{\Delta\text{Cenp-C}}$–Cenp-A$^{\text{Nuc}}$ complexes.** Corresponding size-exclusion chromatograms are shown in Fig. 4b and Extended Data Fig. 9a. **a, b,** Mutating the Cenp-N DNA-binding groove did not impair CCAN$^{\Delta\text{Cenp-C}}$ assembly. **c,** Wild-type CCAN$^{\Delta\text{Cenp-C}}$ forms a complex with Cenp-A$^{\text{Nuc}}$. **d,** Mutating the Cenp-N DNA-binding groove disrupts CCAN$^{\Delta\text{Cenp-C}}$–Cenp-A$^{\text{Nuc}}$ interactions. **e,** Mutating the L1 loop of Cenp-A did not destabilize CCAN$^{\Delta\text{Cenp-C}}$–

Cenp-A$^{\text{Nuc}}$ interactions. **f,** Deletion of the N terminus of Cenp-A (1–129) ($^{\Delta\text{N}}$Cenp-A$^{\text{Nuc}}$) did not impair CCAN$^{\Delta\text{Cenp-C}}$–Cenp-A$^{\text{Nuc}}$ interactions. **h,** Both CCAN$^{\Delta\text{Cenp-C}}$ and CCAN$^{\Delta\text{Cenp-C}}$–Cenp-N$^{\text{Mut}}$ bound poorly to H3$^{\text{N}}$–Cenp-A$^{\text{Nuc}}$. The experiments shown were performed independently in triplicate with similar results. For gel source data, see Supplementary Fig. 1.
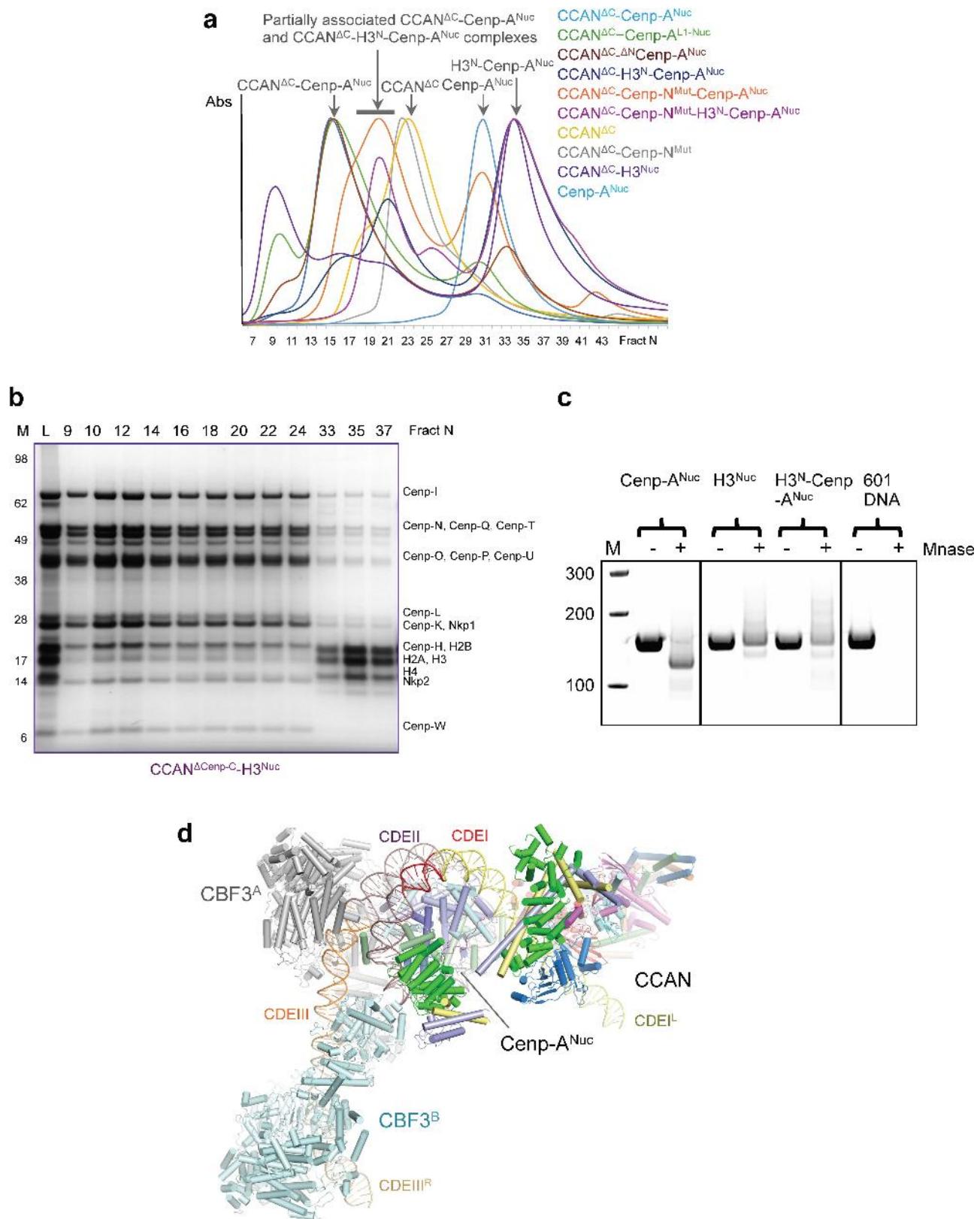
**Extended Data Fig. 9** | See next page for caption.

**Extended Data Fig. 9 | Testing of CCAN$^{\Delta Cenp-C}$ binding to Cenp-A$^{Nuc}$.**
**a**, Comparative SEC profiles (Agilent Bio SEC-5 column) for wild-type CCAN$^{\Delta Cenp-C}$ and the Cenp-N$^{Mut}$ of CCAN$^{\Delta Cenp-C}$ to Cenp-A$^{Nuc}$ and its modifications (Cenp-A$^{Nuc}$-L1$^{Nuc}$, $^{\Delta N}$Cenp-A$^{Nuc}$ and H3$^N$-Cenp-A$^{Nuc}$) and H3$^{Nuc}$. Mutating the L1 loop (Cenp-A$^{L1-Nuc}$) of Cenp-A or deletion of the N-terminal 129 residues ($^{\Delta N}$Cenp-A$^{Nuc}$) did not destabilize CCAN$^{\Delta Cenp-C}$–Cenp-A$^{Nuc}$ interactions. By contrast, CCAN with the Cenp-N$^{Mut}$ bound less well and both CCAN and CCAN–Cenp-N$^{Mut}$ hardly bound to H3$^N$–Cenp-A$^{Nuc}$. (CCAN$^{\Delta C}$, CCAN$^{\Delta Cenp-C}$). Associated SDS–PAGE is shown in Extended Data Figs. 8, 9b). **b**, Coomassie-blue-stained

SDS–PAGE showed that CCAN$^{\Delta Cenp-C}$ did not associate with H3$^{Nuc}$.
**c**, Micrococcal nuclease digestion of Cenp-A$^{Nuc}$, H3$^{Nuc}$ and H3$^N$–Cenp-A$^{Nuc}$. Widom 601 DNA is shown as a control. The H3$^{Nuc}$ and H3$^N$–Cenp-A$^{Nuc}$ protect a similar and longer length of DNA compared with Cenp-A$^{Nuc}$. **d**, Model of CBF3[60] bound to CCAN–Cenp-A$^{Nuc}$, indicating that CBF3 would not associate with a fully assembled kinetochore, consistent with proteomic data[62]. The experiments shown in **a**–**c** were performed independently in triplicate with similar results. For gel source data, see Supplementary Fig. 1.

**a**

1.23 MDa

Cross-linked *S. cerevisiae* CCAN–Cenp-A^Nuc

**b**

1.38 MDa    526 kDa

Uncross-linked *S. cerevisiae* CCAN–Cenp-A^Nuc

**c**

839 kDa    650 kDa

*S. cerevisiae* CCAN alone

**d**

**e**

**f**

**g**

**h**

Observed 2D classes    Calculated 2D classes

**i**

microtubule

Outer-KT

DNA-binding surface    Cenp-LN    Cenp-A^Nuc

CCAN    Cenp-HIK^Head_-TW    Cenp-OPQU+

**j**

microtubule

Outer-KT

Cenp-OPQU+    Cenp-HIK

CCAN    Cenp-LN

Cenp-HIK^Head_-TW    Cenp-A^Nuc

Cenp-T link to outer KT
Cenp-U link to outer KT
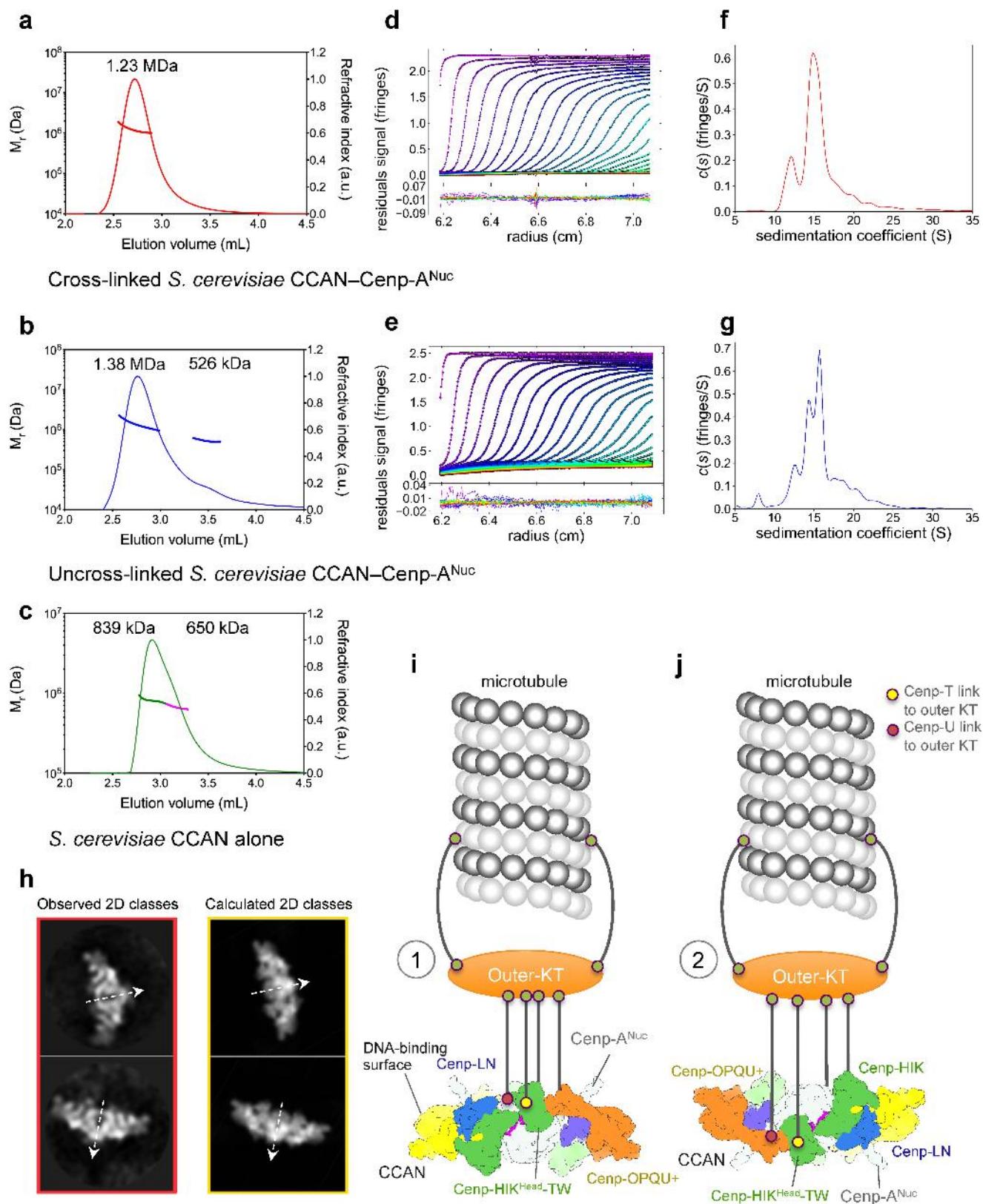
**Extended Data Fig. 10 |** See next page for caption.

**Extended Data Fig. 10 | *S. cerevisiae* CCAN–Cenp-A$^{Nuc}$ comprises two CCAN complexes in solution. a–c,** The predicted mass of (CCAN)$_2$–Cenp-A$^{Nuc}$ is 1.31 MDa, (CCAN)$_1$–Cenp-A$^{Nuc}$ is 0.77 MDa and that of a CCAN dimer 1.09 MDa (Extended Data Table 2). Representative SEC–MALS data for crosslinked *S. cerevisiae* CCAN–Cenp-A$^{Nuc}$ complex (**a**), run independently in triplicate with similar results, average molecular mass is 1.23 MDa ((CCAN)$_2$–Cenp-A$^{Nuc}$); uncrosslinked *S. cerevisiae* CCAN–Cenp-A$^{Nuc}$ complex (**b**), run independently in triplicate with similar results, with average masses of 1.38 MDa ((CCAN)$_2$–Cenp-A$^{Nuc}$) and 526 kDa (CCAN); and *S. cerevisiae* CCAN alone (**c**), run independently in duplicate with similar results, with average masses of 839 kDa for the leading edge (green) and 650 kDa for the trailing edge (magenta), suggesting a non-resolved monomer–dimer equilibrium. **d, e,** Velocity analytical ultracentrifugation of crosslinked (**d**) and uncrosslinked (**e**) *S. cerevisiae* CCAN–Cenp-A$^{Nuc}$ complexes with residuals to the fits shown in **f** and **g. f, g,** Fit of a $c(s)$ distribution model for the crosslinked complex (**f**), the major species sediments at 15.8S ($S_{w,20} = 26.1S$) with a minor species at 12.1S ($S_{w,20} = 20.0S$) that corresponds to calculated masses of 1.34 MDa ((CCAN)$_2$–Cenp-A$^{Nuc}$) and 896 kDa (possibly (CCAN)$_1$–Cenp-A$^{Nuc}$), respectively, with a fitted value of 1.761 for the frictional ratio. **g,** Fit for uncrosslinked samples, the major species is resolved into two species that sediment at 14.3S ($S_{w,20} = 22.6S$) and 15.7S ($S_{w,20} = 24.9S$) with a minor species at 12.3S ($S_{w,20} = 19.4$ S),

which gave masses of 1.32 MDa ((CCAN)$_2$–Cenp-A$^{Nuc}$) and 1.15 MDa ((CCAN)$_2$) for the major species and 716 kDa ((CCAN)$_1$–Cenp-A$^{Nuc}$) for the minor species. The experiments shown in **d–g** were performed independently in triplicate with similar results. **h,** Examples of two 2D class averages showing the (CCAN)$_2$–Cenp-A$^{Nuc}$ particles viewed in the plane of the C2 symmetry axis (red outline) (data from Extended Data Fig. 2c) and the 2D reprojections of a modelled (CCAN)$_2$–Cenp-A$^{Nuc}$ based on the CCAN–Cenp-A$^{Nuc}$ cryo-EM reconstruction (yellow outline) (shown in **i**). There is a close correspondence in shape and dimensions between the calculated reprojections and the observed 2D classes. The two-fold symmetry axes of the (CCAN)$_2$–Cenp-A$^{Nuc}$ complex are shown as dashed arrows. **i, j,** Two alternative models for how CCAN assembled onto a Cenp-A nucleosome would interact with the outer kinetochore–microtubule interface (Supplementary Video 2). **i,** In scenario (1), CCAN interacts with the outer kinetochore from the same side as the DNA-binding surface. Microtubules attached to the outer kinetochore would hoist CCAN from below the over-lying nucleosome and out-stretched DNA. **j,** In scenario (2), the microtubule-outer kinetochore interface contacts CCAN from the opposite side to the CCAN DNA-binding surface. Outer-kinetochore (outer-KT): KMN network and microtubule-attachment complexes, Dam1–DASH (budding yeast) and Ska proteins of vertebrates. The combined dimension of (CCAN)$_2$–Cenp-A$^{Nuc}$ (32 nm) matches that of the hub at the centre of the yeast kinetochore[63].

**Extended Data Table 1 | Cryo-EM data collection, refinement and validation statistics**

| | CCAN (EMDB-4580) (PDB 6QLE) | CCAN–Cenp-A$^{Nuc}$ (EMDB-4579) (PDB 6QLD) | Mask1 (EMDB-4581) (PDB 6QLF) | Mask2 (EMDB-4971) |
|---|---|---|---|---|
| **Data collection and processing** | | | | |
| Magnification | 75,000 | 75,000 | 75,000 | 75,000 |
| Voltage (kV) | 300 | 300 | 300 | 300 |
| Electron exposure (e–/Å$^2$) | 32 | 32 | 32 | 32 |
| Defocus range (µm) | 2.0-2.8 | 2.0-2.8 | 2.0-2.8 | 2.0-2.8 |
| Pixel size (Å) | 1.09 | 1.09 | 1.09 | 1.09 |
| Symmetry imposed | C1 | C1 | C1 | C1 |
| Initial particle images (no.) | 1,796,016 | 1,796,016 | 1,796,016 | 1,796,016 |
| Final particle images (no.) | 618,459 | 193,882 | 618,459 | 618,459 |
| Map resolution (Å) | 3.55 | 4.15 | 3.45 | 3.83 |
| FSC threshold | 0.143 | 0.143 | 0.143 | 0.143 |
| Map resolution range (Å) | 3.0-5.5 | 3.5-7.0 | 3.0-5.5 | 3.0-5.5 |
| | | | | |
| **Refinement** | | | | |
| Initial model used (PDB code) | 5MU3, 6EQT, 4JE3, 5W94 | 3AN2, 4X23, 5MU3, 6EQT, 4JE3, 5W94 | 5MU3, 6EQT, 4JE3, 5W94 | 5MU3, 6EQT, 4JE3, 5W94 |
| Model resolution (Å) 0.143 FSC threshold | 3.5 | 4.0 | 3.3 | - |
| Model resolution range (Å) | 50 - 3.0 | 50 - 3.6 | 50 - 3.0 | - |
| Map sharpening $B$ factor (Å$^2$) | -139 | -108 | -135 | -172 |
| Model composition | | | | |
| Non-hydrogen atoms | 18,058 | 29,183 | 13,541 | - |
| Protein residues | 2,401 | 3,172 | 1,790 | - |
| Ligands | 0 | 248 | 0 | - |
| $B$ factors (Å$^2$) | | | | |
| Protein | 78.6 | 82.2 | 67.2 | - |
| Ligand | - | 245.8 | - | - |
| R.m.s. deviations | | | | |
| Bond lengths (Å) | 0.004 | 0.004 | 0.005 | - |
| Bond angles (°) | 0.798 | 0.793 | 0.828 | - |
| Validation | | | | |
| MolProbity score | 1.39 | 1.57 | 1.45 | - |
| Clashscore | 2.78 | 4.80 | 2.99 | - |
| Poor rotamers (%) | 0.11 | 0.08 | 0.19 | - |
| Ramachandran plot | | | | |
| Favored (%) | 95.30 | 94.78 | 94.76 | - |
| Allowed (%) | 4.60 | 5.02 | 5.04 | - |
| Disallowed (%) | 0.10 | 0.20 | 0.20 | - |

**Extended Data Table 2 | Table of CCAN subunits**

| Subunit | *S.c.* name | Length | Mol. Mass kDa | Domain/Region 1 | Domain/ Region 2 | Domain/ Region 3 | Disordered regions | Sequence built as polyA |
|---|---|---|---|---|---|---|---|---|
| *Sc*Cenp-A nucleosome | | | | | | | | |
| **Cenp-A** | Cse4 | 229 | 26.8 | α-helix and disordered 1-131 | Histone fold 132-229 PDB 3AN2 *Hs* Cenp-A | - | 1-111,131-136,227-229 | 112-130 |
| **H2A** | | 132 | 14.0 | Histone fold PDB 1ID3 *Sc* H2A | - | - | - | - |
| **H2B** | | 132 | 14.2 | Histone fold PDB 1ID3 *Sc* H2A | - | - | - | - |
| **H4** | | 103 | 11.4 | Histone fold PDB 1ID3 *Sc* H2A | - | - | - | - |
| **601 DNA** | | 147 bp | 90.6 | | | | | |
| | | | | | | | | |
| **Cenp-C** | Mif2 | 549 | 62.5 | Cenp-C motif 283-304 PDB 4X23 | Cupin fold 365-530 | - | 1-283,306-549 | - |
| | | | | | | | | |
| Cenp-HIK−TW complex (*Sc*Ctf3 complex + Cenp-TW) | | | | | | | | |
| **Cenp-H** | Mcm16 | 181 | 21.1 | α-helix: *De novo* 4-136 | α-helix: PDB 5Z07 *Ct* Cenp-I 143-181 | - | 1-3,41-44,75-78,137-142 | - |
| **Cenp-I** | Ctf3 | 733 | 84.3 | Heat repeats PDB 5Z07 *Ct* Cenp-I 5-241 | Heat repeats: *De novo* | - | 242-332,526-531,597-601,620-624,657-663,677-689 | 321-330,664-676 |
| **Cenp-K** | Mcm22 | 239 | 27.6 | α-helix: *De novo* 7-128 | α-helix: PDB 5Z07 *Ct* Cenp-I 143-236 | - | 1-6,42-49,61-68,129-142 | - |
| **Cenp-T** | Cnn1 | 361 | 41.3 | Histone fold | | | ND | ND |
| **Cenp-W** | Wip1 | 98 | 10.2 | Histone fold | | | ND | ND |
| | | | | | | | | |
| **Cenp-LN complex** | | | | | | | | |
| **Cenp-L** | Iml3 | 245 | 28.0 | α/β fold PDB 4JE3 *Sc* Cenp-L | - | - | - | - |
| **Cenp-N** | Chl4 | 458 | 52.7 | Pyrin (1-102) Cenp-N fold (103-262) PDB 6EQT *Hs* Cenp-N | Cenp-N linker domain *de novo* (262-373) | Dimerization (375-468) PDB 4JE3 *Sc* Cenp-N | 1-4,47-50,166-192,310-316,338-373,452-458 | - |
| | | | | | | | | |
| Cenp-OPQU+ complex (*Sc*COMA+ complex) | | | | | | | | |
| **Cenp-O** | Mcm21 | 368 | 43.0 | RWD PDB 5MU3 *Kl* Ctf19 | - | - | 1-152,332-338 | - |
| **Cenp-P** | Ctf19 | 369 | 42.8 | RWD PDB 5MU3 *Kl* Ctf19 | - | - | 1-96,111-123,286-292,308-313 | 97-110 |
| **Cenp-Q** | Okp1 | 406 | 47.4 | α-helix: De novo | - | - | 1-160,220-228,304-319,392-406 | 161-219 |
| **Cenp-U** | Ame1 | 324 | 37.5 | α-helix: De novo | - | - | 1-130,157-165,267-276 | 131-156 |
| **Nkp1** | Nkp1 | 238 | 27.0 | α-helix: *De novo* | - | - | 1,124-135 | 24-32,217-238 |
| **Nkp2** | Nkp2 | 153 | 17.9 | α-helix: *De novo* | - | - | 1-2,25-35 | 133-153 |

Details of structured regions of CCAN subunits built into the cryo-EM density maps are indicated, including regions built as polyAla. The calculated molecular masses for CCAN and Cenp-A$^{Nuc}$ complexes are (i) CCAN: 543.3 kDa, (ii) CCAN dimer: 1.09 MDa, (iii) Cenp-A$^{Nuc}$: 223 kDa, (iv) (CCAN)$_1$−Cenp-A$^{Nuc}$: 0.766 MDa and (v) (CCAN)$_2$−Cenp-A$^{Nuc}$ 1.31 MDa.

# nature research

Corresponding author(s):   David Barford

Last updated by author(s):   Aug 22, 2019

# Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

## Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

| n/a | Confirmed | |
|---|---|---|
| ☒ | ☐ | The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement |
| ☒ | ☐ | A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly |
| ☒ | ☐ | The statistical test(s) used AND whether they are one- or two-sided<br>*Only common tests should be described solely by name; describe more complex techniques in the Methods section.* |
| ☒ | ☐ | A description of all covariates tested |
| ☒ | ☐ | A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons |
| ☒ | ☐ | A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals) |
| ☒ | ☐ | For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted<br>*Give P values as exact values whenever suitable.* |
| ☒ | ☐ | For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings |
| ☒ | ☐ | For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes |
| ☒ | ☐ | Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated |

*Our web collection on statistics for biologists contains articles on many of the points above.*

## Software and code

Policy information about availability of computer code

| | |
|---|---|
| Data collection | Commercial software: EPU from Thermo Fisher Scientific was used for automated cryo-EM data collection. |
| Data analysis | Cryo-EM data were analyzed using the software MotionCor2 (version 2.1), GCTF (version 0.5), RELION2.1 (version 2.1), RELION3.0 (version 3.0), SIMPLEPRIME (version ) and RESMAP (version 1.1.4). Model building and refinement were performed using COOT (version 0.8.9.2) and Phenix (version 1.15.2) and validated in COOT (version 0.8.9.2) and MolProbity (version 4.2). Visualization was performed with COOT (version 0.8.9.2), PyMOL (version 1.8.4.1, Chimera (version 1.8.1) and ChimeraX (version 0.8). Structure figures were generated use PyMOL (v1.8.4.1) and Chimera (version 1.8.1).<br>Sequence alignments were performed and displayed with JALVIEW (version 1.0).<br>Structure prediction was performed with the PHYRE2 web tool.<br>Protein secondary structure and disordered regions were predicted with the PHYRE2 and PSIPred web tools.<br>AUC data were analysed in SEDFIT v16.1 and Sednterp (version 20130813 beta). SEC-MALS data were analysed using ASTRA version 5.3.4.20 software (Wyatt Technologies) and data were plotted with the program PRISM (version 8.2.0) (GraphPad Software Inc.).<br>XL-MS data were analysed using Proteome Discoverer 2.3 (version 2.3.0.522) and the xVis web tool. |

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

## Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

EM maps are deposited with EMDB with accession codes EMD-4580 (CCAN), EMD-4579 (CCAN-Cenp-ANuc), EMD-4581 (Mask1) and EMD-4971 (Mask2). Protein coordinates are deposited with RCSB with accession codes 6QLE (CCAN), 6QLD (CCAN-Cenp-ANuc) and 6QLF (Mask1). The cross-linking mass spectrometry raw files, the associated output and databases are deposited through the ProteomeXchange Consortium 48 via the PRIDE partner repository with the dataset identifier PXD013769.

# Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences    ☐ Behavioural & social sciences    ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](#)

# Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

| | |
|---|---|
| Sample size | We collected 9002 cryo-EM images for the CCAN-Cenp-A dataset and 910 cryo-EM images for the Cenp-HIK dataset. The total number of particles for the CCAN-Cenp-A dataset was 1,796,061 and that for the Cenp-HIK dataset was 123,215. Sample sizes were estimated on the basis of previous studies using similar methods and analyses that are widely published. |
| Data exclusions | No data were excluded from the analysis. |
| Replication | All attempts at replication were successful and reproducible. At least three independent biological repeats per experiment where representative data are shown. Structure determination does not require replication. |
| Randomization | Samples were not allocated into groups. Randomization is not relevant to this study. |
| Blinding | Blinding was not relevant to this study because there was no group allocation. |

# Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

### Materials & experimental systems

| n/a | Involved in the study |
|---|---|
| ☐ | ☒ Antibodies |
| ☐ | ☒ Eukaryotic cell lines |
| ☒ | ☐ Palaeontology |
| ☒ | ☐ Animals and other organisms |
| ☒ | ☐ Human research participants |
| ☒ | ☐ Clinical data |

### Methods

| n/a | Involved in the study |
|---|---|
| ☒ | ☐ ChIP-seq |
| ☒ | ☐ Flow cytometry |
| ☒ | ☐ MRI-based neuroimaging |

## Antibodies

| | |
|---|---|
| Antibodies used | One primary antibody used: MOUSE ANTI STREP-TAG CLASSIC:HRP. Anti-Strep antibody (Source: Bio-Rad, Catalogue code:MCA2489P, Batch No: 147517). Dilution 1 to 1000. |
| Validation | Mouse anti Strep-Tag Classic antibody, clone Strep-tag II , also known as StrepMAB-Classic, recognizes Strep-tag II, a widely used tag in protein expression applications. This antibody recognizes both-C- and N-terminal Strep-tag II and is especially suited to Western blot applications.<br><br>Validation: HCA182 specificity ELISA using various antigens (A: Human Serum, B: human IgG1/kappa from myeloma plasma, C: Rituximab, D: Ustekinumab, E: Infliximab, F: Adalimumab, G: Alemtuzumab and H: Bevacizumab) as coating components |

followed by Human anti Avastin®(HCA182) and HRP conjugated Mouse anti Strep-tag (MCA2489P) as detection reagent.

Literature on Bio-Rad web-site:

1. Renzi F et al. (2015) Glycan-Foraging Systems Reveal the Adaptation of Capnocytophaga canimorsus to the Dog Mouth. MBio. 6 (2): .pii: e02507-14.
2. Gordon, C.A. et al. (2015) NUSAP1 expression is upregulated by loss of RB1 in prostate cancer cells. Prostate. 75 (5): 517-26.
3. Mavrakis, M. et al. (2016) Purification of recombinant human and Drosophila septin hexamers for TIRF assays of actin-septin filament assembly. Methods Cell Biol. 136: 199-220.
4. Oda, S. et al. (2015) Crystal Structure of Marburg Virus VP40 Reveals a Broad, Basic Patch for Matrix Assembly and a Requirement of the N-Terminal Domain for Immunosuppression. J Virol. 90 (4): 1839-48.
5. Renzi, F. et al. (2015) Glycan-foraging systems reveal the adaptation of Capnocytophaga canimorsus to the dog mouth. MBio. 6 (2): e02507.

# Eukaryotic cell lines

Policy information about cell lines

| Cell line source(s) | High-5 insect cell: Trichoplusia ni: expression system.<br><br>S. cerevisiae strains:<br>1. AEY4992: with a chl4 deletion and cse4-R37A mutation (chl4Δ cse4-R37A), (MATα ade2-101 lys2 his3-11,15 trp1-1 leu2-3,112 ura3-1 can1-100 chl4Δ::kanMX Cse4R37A).<br>2. W303 (wild type strain: MATα ade2-101 his3-11,15 trp1-1 leu2-3,112 ura3-1). |
|---|---|
| Authentication | The High-5 insect cell line was not authenticated. The S. cerevisiae strains were authenticated refs 27, 41. |
| Mycoplasma contamination | The High-5 insect cell line was not tested for mycoplasma contamination. Yeast strains do not have mycoplasma and were not tested for mycoplasma contamination |
| Commonly misidentified lines (See ICLAC register) | None |

# CORRECTIONS & AMENDMENTS

## Author Correction: A rigorous electrochemical ammonia synthesis protocol with quantitative isotope measurements

Suzanne Z. Andersen, Viktor Čolić, Sungeun Yang,
Jay A. Schwalbe, Adam C. Nielander, Joshua M. McEnaney,
Kasper Enemark-Rasmussen, Jon G. Baker,
Aayush R. Singh, Brian A. Rohr, Michael J. Statt,
Sarah J. Blair, Stefano Mezzavilla, Jakob Kibsgaard,
Peter C. K. Vesborg, Matteo Cargnello, Stacey F. Bent,
Thomas F. Jaramillo, Ifan E. L. Stephens, Jens K. Nørskov &
Ib Chorkendorff

In this Letter, the $y$-axis label of Fig. 4b should read 'Yield of $NH_3$ ($\mu$mol)' rather than 'Yield of $NH_3$ (mmol)'. The original paper has been corrected online.

# CORRECTIONS & AMENDMENTS

## Author Correction: Global analysis of streamflow response to forest management

Jaivime Evaristo & Jeffrey J. McDonnell

In this Article, the authors declared no competing interests; however, in the interests of transparency, the authors wish to amend the Competing Interests statement to read: 'J.J.M. provided consulting advice to Arauco Chile on three occasions, most recently in 2015.'. The original Article has been corrected online.

# CORRECTIONS & AMENDMENTS

## Editorial Expression of Concern: Global analysis of streamflow response to forest management

Jaivime Evaristo & Jeffrey J. McDonnell

The editors of *Nature* have become aware that this Article contains at least two serious technical errors. First, the assembled dataset of paired watershed studies, used to assess the streamflow response to forest removal and planting, contains errors in the percentage change in streamflow associated with land cover modifications. Second, the effects of continent-wide forest removal on streamflow (shown in Table 1) are overestimated, because the authors assumed a starting condition of 100% forest cover. We are aware that other technical concerns have also been raised; we are investigating these critiques and will provide an update once a resolution has been reached. The authors have been informed of this Editorial Expression of Concern, and are in agreement.

# CAREERS
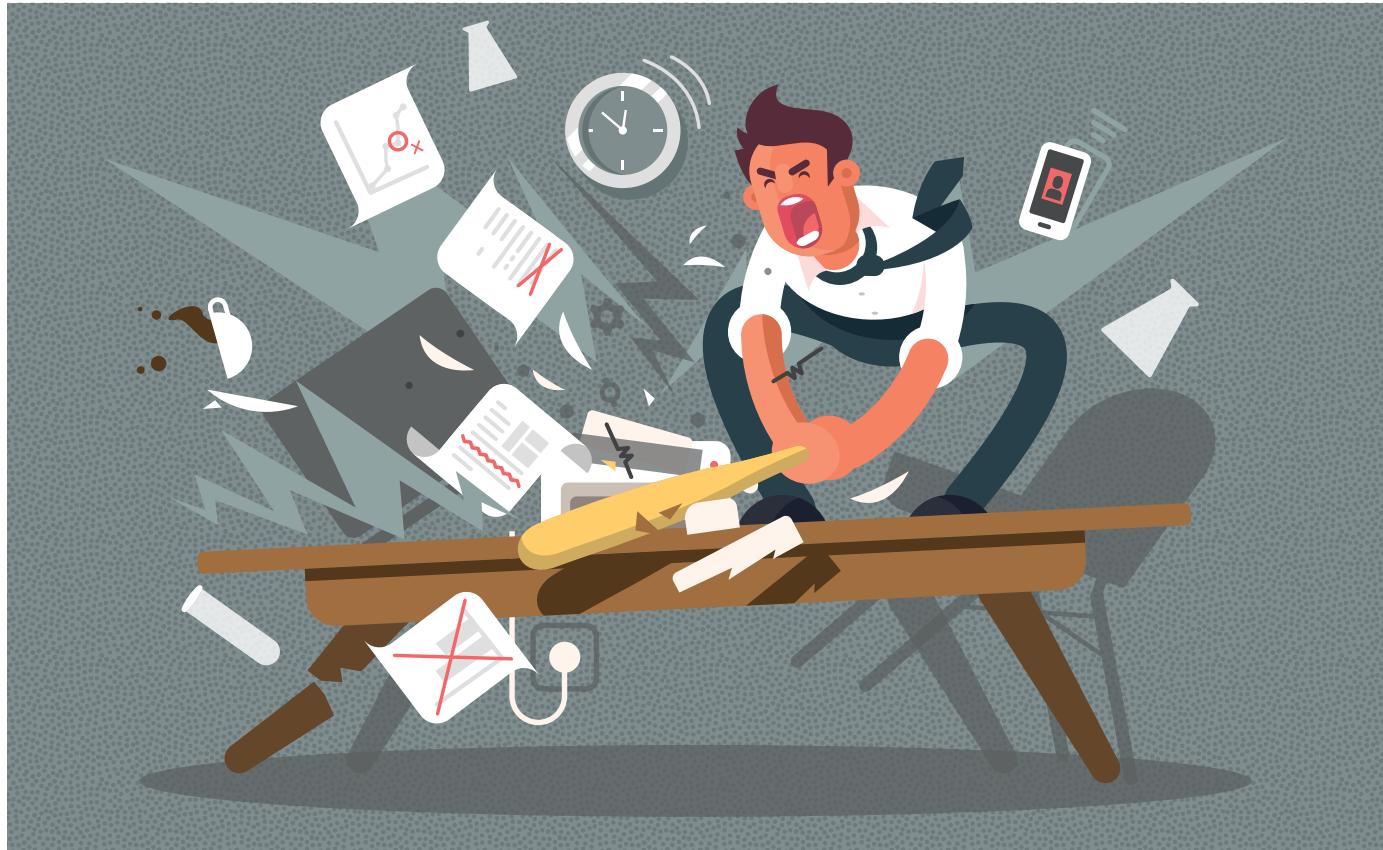
ORGANIZATIONAL SKILLS

# Avoid PhD deadline rage

*Tips to skip the last-minute panic and take the stress out of submitting your thesis.*

**BY NIC FLEMING**

Horror stories about the final weeks, days and hours before a thesis submission deadline are common among people with PhDs in both the sciences and humanities.

Some are undone by losing their precious words to unresponsive hard drives. Others see their graphs and references mangled by software that can't cope. There are sleep-deprived administrative blunders, formatting problems, severe cases of writer's block and stress-induced disasters. In fact, candidates for whom thesis submission goes entirely to plan are almost certainly in the minority.

*Nature* spoke to individuals who have been through disasters, or have helped others to overcome them, to find tips to get you through submission day.

## PLAN FAR AHEAD

Last August, Mark Bennett was waiting anxiously outside the university print shop, USB stick in hand, when it opened its doors at 9 a.m.. The previous evening, Bennett had ordered three copies of his thesis on the shop's website, and received an e-mail telling him when he could pick them up. But the site hadn't prompted him to upload the document, so he knew something had gone wrong. By that time it was too late to call the printers, and his final deadline was just days away.

Bennett had started his English literature PhD on eighteenth-century travel writing and its relationship to popular fiction at the University of Glamorgan in Pontypridd, UK, in 2008. But through a combination of funding issues, starting a family, following his supervisor's move to the University of Sheffield, UK, and beginning a full-time job, Bennett did not complete his thesis until the end of August 2018. Submitting a day late could have resulted in a fail. "I was in a panic, thinking I'd now have go to an appeal at which it was going to be a 'dog ate my homework scenario', which is really not appropriate at PhD level," says Bennett.

"I've never heard of a PhD student who hasn't had something unexpected or untoward happen, especially in the later stages," says Inger ▶

Mewburn, director of research training at the Australian National University in Canberra. "In 15 years of working with PhD candidates, I've never heard anyone say, 'It was totally fine.'"

As the print shop opened its doors for business, Bennett made his way inside, silently kicking himself at the thought that leaving this seemingly simple task to the last minute might result in him failing the PhD he had started a decade earlier.

The shop assistant who checked the file told Bennett he could not print it because it was a Microsoft Word document. Bennett's sense of dread was exacerbated as their combined efforts to convert it into a PDF using freeware failed. The problem was finally solved by another assistant, who pointed out that it could be converted into a PDF within Word. "It's natural that people want to take every last bit of time to work on their thesis, but they shouldn't assume printing and binding will be a formality," says Bennett, who works at FindAUniversity, a Sheffield-based company that operates several websites for students seeking postgraduate opportunities. "It's worth getting it done well ahead of the deadline."

### BACK UP YOUR DATA IN MULTIPLE PLACES

Although few people would want to return to writing PhDs on typewriters, storing data and text in digital form is not without its own risks. Physicist Leonor Sierra knows this better than most. In 2006 she was most of the way through her PhD on quantum transport in carbon nanotubes at the University of Cambridge, UK, when a number of computers, including hers, were quarantined because of a computer virus. This led to a delay of only a fortnight or so, which might in other circumstances have just been a minor setback; however, her progress had already been slowed by the head of her laboratory moving away and the lab's fabrication facilities being shut down for several months.

Early the following year, Sierra had written almost half of her thesis when the external hard drive she was using to back up her work suddenly stopped working. She was not overly concerned, because her work was also stored on her computer and on CDs. Sierra resolved to get a new hard drive. A week later, however, she tried to turn on her computer only to find it unresponsive. She tried several times to no avail, and then burst into tears.

A computer-scientist friend removed the computer's hard drive, put it into another machine and retrieved all but about a chapter's worth of work. "At the time, it seemed like the end of the world," says Sierra. "But re-writing it didn't take long because I already knew what I wanted to say, and the second version was better, so it was a blessing in disguise."

Since Sierra submitted her PhD in 2007, the rise of cloud-based storage has meant fewer students lose work to hardware failures. That does not, however, mean that digital risks are a thing of the past. "I would advise people to use more than one back-up system, to make use of the cloud, and not to discard early data, printed

---

**CASE STUDY**

## *How to avoid an administrative nightmare*

Margin sizes, forms and printing ink might be the last thing on your mind as deadline day approaches. But leaving administrative requirements to the last minute could be costly.

PhD coach James Hayton advises making a checklist of the following:
- Triple check your deadline from an official source.
- Find out which office you need to hand your thesis to. When does it close?
- What forms do you need to fill in? Who needs to sign them?
- Make note of the required margin size, line spacing and typefaces.
- If your thesis needs binding, what are the specifics? Where can you do it?
- How many copies do you need to submit? It's usually at least two, sometimes more.
- Do you have access to a printer with enough paper and ink, and a back-up?
- Figures can look different when printed, especially in colour. Do early test runs.
- Get someone to check the title page: misspelling your name won't impress.
- Allow time to solve problems caused by compiling separate chapters into one file and format conversion. **N.F.**

---

drafts or other material until the very end," adds Sierra, who now lives in Athens, Georgia, and works as a freelance science writer and editor.

### PROJECT MANAGE YOUR MONSTERS

Many students' struggles to complete their theses are rooted in the organizational difficulties they faced at the start of their PhD programme. Whereas undergraduates are largely expected to learn and understand existing material, there are no answers at the back of the book for PhDs. Supervisors offer directions, but candidates must draw their own maps as they go along. This means they must manage their own schedules.

Project-management skills are therefore key, says Sara Shinton, head of researcher development at the University of Edinburgh, UK. At that institution's induction events, candidates receive a wall chart with 48 empty boxes representing months, which they are encouraged to fill with important events, plans and deadlines relating to their PhDs. The idea is that students will find writing a thesis easier if they keep it in mind as they plan and complete earlier aspects of the programme, such as reviewing the literature, attending conferences, doing placements, devising experiments and collecting results. "If you're reflecting on the bigger questions

---

through the process, then you'll be in a much better position to weave the narrative when it comes to the end," says Shinton.

Small formatting and referencing issues can grow into substantial problems in the final days before submission. Some PhD students get into trouble by leaving details such as references, fonts, text size and graph format until later on, says James Hayton, a PhD coach and author of the 2015 book *PhD: An Uncommon Guide to Research, Writing and PhD Life*. "A common trap is leaving those awkward little things to the end, and it taking longer than expected," he says. "I advise choosing any referencing software, sorting out things like formatting and graphs, and getting things as close to submittable as possible early on."

### GET WRITING

Many struggle with the writing process itself. Having a clear timeline of when you will complete drafts of chapters can help to keep you from falling behind. Mewburn, who runs a blog called the The Thesis Whisperer along with three-day thesis boot camps for PhD candidates at the Australian National University, advises students to finish their first full draft six months ahead of the deadline. "People say to me 'no way', and then I go through all the practical details that can go wrong with things like getting supervisor sign-off, using the wrong template, finding a decent copy editor, and dealing with their input, as well as all the normal life problems which will become magnified and harder to deal with."

Mewburn says that all her boot-camp attendees write at least 5,000 words over the 3 days, and some hit 20,000. As well as using motivational techniques such as awarding different-coloured, giant Lego blocks as prizes for hitting various targets, she also teaches generative writing, a technique designed to get the words flowing by, for example, advising writers to suppress the desire to self-edit as they type. (To do this, Mewburn covers the delete keys on boot-camp-participants' keyboards with fuzzy stickers to prevent attendees from auto-editing.)

Rowena Murray, director of research at the University of the West of Scotland near Glasgow, UK, advises PhD candidates to write a 750-word summary of their thesis within eight months of their deadline. "It makes them focus on what will go into each chapter, the coherence of the structure and the macro arguments, as opposed to just the micro details they are enmeshed in at that time."

### CONTROL THE CONTROLLABLE

Universities and their departments each have specific administrative requirements for thesis submission, and PhD candidates can reduce the risk of last-minute headaches by getting to grips with these criteria early on — and possibly committing them to paper, says Hayton (see 'How to avoid an administrative nightmare'). "When you are stressed, it's best not to rely on your short-term memory," he says. "It's better to make a checklist of the required paperwork."

Some thesis-submission complications are beyond the powers of even the most organized students to do anything about. If your lab burns down, taking your experiment and results with it, no amount of planning or preparation will help. However, examiners are not looking to fail candidates, and will generally take pity on those who have genuinely had bad luck. "There are always the acts of God-type events," says Shinton. "Funders and institutions are always going to look sympathetically at such cases."

Given the hard work and sacrifice required to gain a PhD and the wide variety of things that can go wrong, some might wonder whether it is worth it. The answer will vary on a case-by-case basis, depending partly on individuals' career paths and other goals. Some who advise PhD candidates say it is important to bear in mind the scope for personal development that gaining the prized qualification can bring.

Mewburn, for example, thinks that completing her PhD on the use of hand gestures in the teaching of architecture gave her the confidence to take on a number of complex professional projects. She uses the Finnish word '*sisu*' to describe the grim determination in the face of adversity that individuals must go through to get their PhDs. "The process of doing a PhD shows you what you are capable of," she says. "If it is done well, it can give you an intense sense of achievement and power. "Plus," Mewburn adds, "it's nice when people call you 'doctor' on aeroplanes." ∎

**Nic Fleming** *is a freelance writer based in Bristol, UK.*

# COLUMN
# Teaching is a privilege

Scientists should embrace teaching responsibilities, advises **Sarah A. Gagliano Taliun**.

"A postdoc is just like a faculty position minus all the hassles of teaching," a visiting professor told me and a handful of postdoctoral colleagues during an informal networking lunch earlier this year.

I disagree with this attitude towards education. Teaching at the university level is not and should not be considered a burden or chore that just needs to be done. It is a crucial part of academia, and it is essential that mentors portray it as such. We all want to do scientifically sound research, and, without question, we should all strive to be effective teachers. Through teaching, researchers are responsible for the education of the next generation of scientists ,who will use their own unique ideas and skill sets to advance their fields.

In both my PhD programme and my postdoctoral fellowship, I have sought out teaching opportunities because I see them as an opportunity for enrichment, rather than a hindrance. I have supervised undergraduates during an intensive summer research programme, and have mentored numerous students doing research. Also, as a postdoctoral fellow, I have co-instructed several graduate-level courses. Each time I find myself in a teaching role, I try to do it better.

I work to improve the delivery of the lesson, to induce a deeper level of critical thinking through my exam questions and to incorporate new teaching strategies to meet the needs of a wider range of learners. I learn from my students. Through their fresh perspectives, I am able to rethink my research as well as the current state of the field and where it is going. For example, questions from my students helped me to reconsider the accepted threshold for 'genome-wide significance' and how it might change.

From my experiences, I have three pieces of advice to help researchers become better teachers.


Teaching and research are both integral parts of science.

Approach teaching with an open mind. The predominant attitude in the sciences needs to shift: teaching is not a waste of prized research time. Certainly, there are academics who value the responsibility of teaching, but this group needs to become the majority.

Reach out for support when planning a class. Most of us are not innate teachers, just as most of us are not innate researchers. As with developing any skill, learning to teach is a process that requires trial and error and lots of practice. To this end, many universities offer professional-development programmes designed for graduate students, postdoctoral fellows or faculty members to improve teaching practices and techniques in the classroom, the laboratory and beyond. It is never too early or too late to work on developing these skills, many of which are applicable outside the classroom, such as when mentoring students who are doing research or giving oral presentations.

Prepare thoroughly so that the content and flow of the lesson is concise and coherent, and is tailored to the audience. This preparation takes time, but by doing it you will at the same time develop new ideas on presenting your own research (through verbal, written or visual means) to non-specialists, thus broadening its reach.

I am working towards a career in academia and am aware of the ever-increasing pressures on researchers to publish in high-quality journals, secure funding and present at conferences. Teaching is often lower down on this list of priorities. I feel that science needs to rethink its positioning.

Teaching at the university level should not be seen as a hassle in academia, but rather as a skill to be developed and a responsibility to be taken seriously. Teaching does not have to decrease research productivity — it can greatly enhance research if we allow it to. ∎

**Sarah A. Gagliano Taliun** *is a postdoctoral research fellow in biostatistics at the University of Michigan in Ann Arbor.*

# INFRINGEMENT

*Brought down to Earth.*

BY TIMOTHY J. GAWNE

I opened the front door and was not surprised by the tall man covered in bright blue feathers because I was distracted by the kilometre-wide flying saucer hovering overhead.

The feather-covered man looked down at a clipboard, and said: "Are you Mr Floyd Bromley of Birmingham, Alabama?"

"Yes, I am." I noticed that the street in front of my house was lined with police, and that yellow barricades had been erected all around.

"Well, splendid then," said the man as he looked up from his clipboard. "This shouldn't take more than a few moments of your time. May I come in?"

"Um, sure," I said, and I motioned for him to enter.

"You are", said the man, "wondering what is going on." The man wore a broad red sash covered with complex hieroglyphics that, as I watched, re-formed themselves to spell out 'COPYRIGHT POLICE'.

I pointed to the sash. "Copyright Police?"

"Ah good, the multilingual reformatter is working," said the man. "Yes, I am an official of the universal body that deals with intellectual-property violations. I'm investigating a claim of infringement."

"Oh," I said. "And what is being infringed?"

"*Allegedly* infringed," said the man. "Why, Earth, of course."

"Earth?" I said.

"Indeed," said the man. "Earth is quite a popular product, and allegations of it being pirated are taken quite seriously."

"There is more than one Earth?" I said.

"Certainly," said the man. "At last check there have been over one million sold."

"And what do people do with all of these Earths?" I asked.

"Most use their Earths as decorations or conversation pieces, as I see you have a lovely brass clock on your mantelpiece. Some take on human form and go down onto the surface of the planet itself, to experience it directly. And there are always enthusiasts, who enhance the technology, overclock evolution — there are clubs and competitions for that."

"So many Earths — are they all the same?" I said. "If someone could order an entire planet, wouldn't they want something unique?"

"There's good business in custom planets, but most people are happy with mass-produced items." He gestured at my living room. "After all, most of the furniture and appliances that you have are identical to what can be found in thousands of other homes."
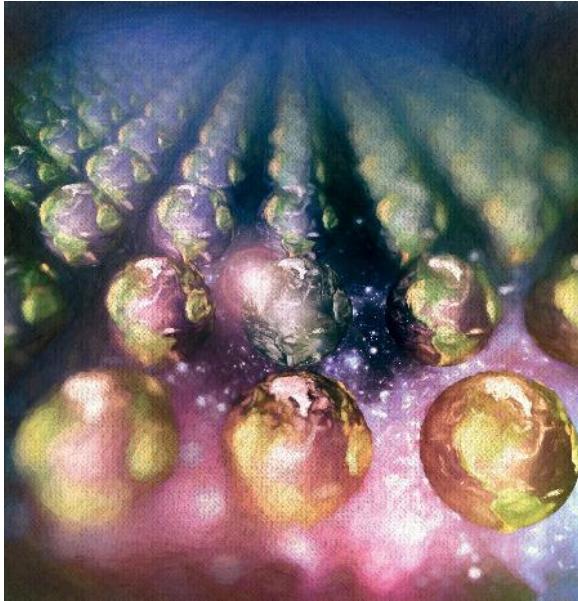
I nodded. "I suppose that makes sense. But what does all this have to do with me?"

The man looked down at his clipboard again. "It has been alleged that this Earth is not a suitably authorized Earth, but is in fact a pirated copy. I have been sent here to determine the truth of said allegation."

The door opened and another man entered. This one was not covered in blue feathers, but was wearing a striped short-sleeved shirt with khaki trousers and tennis shoes. It took me a moment to realize that the other person was, in fact, me. I've obviously seen myself in mirrors and photographs, but I've never seen myself in person. It was strange.

"This", said the feathered man, "is the Floyd Bromley from the manufacturer's standard reference Earth. I'm going to use him to certify my calibrations, then take a few measurements on you to check."

A complex arrangement of brass spheres and rings materialized in front of the feathered man. He took it and moved it around the other me. Then he started to move it around me.

"Hmm … blood cells check out," said the feathered man. "But the muscle fibres are a hack, and the brain — good lord, what a mess they've made of the thalamus, and the cerebellum is hardly any better. And the mitochondria are rubbish. I'm sorry, Mr Bromley, but this Earth is in fact a pirated copy — and a poor one at that."

Floyd Standard shook his head. "Tough break …" I noticed that this other me was taller, and his skin was smoother. I wear glasses, and he did not. I have grey hair, and his was thick and black. Was I really a cheap copy?

"What happens now?" I asked.

"Well," said the feathered man, "the counterfeit item will be impounded and destroyed. The offending party will be heavily fined …"

"It might just be a cheap copy," I said, "but I'm living on it. You'd snuff out billions of sentient life forms over a case of copyright violation?"

The feathered man wrinkled his nose. "*Partial* sentients. Although I do sympathize. You could appeal to the universal council."

"I could?" I said. "How?"

"I can do it for you," said the feathered man, "no trouble at all." His eyes defocused and he mumbled something unintelligible to himself. Then, after barely ten seconds had passed, he refocused his eyes.

"Sorry, the council has rejected your appeal," said the man.

"That's it? Ten seconds, and it's rejected?"

"I will have you know," said the feathered man, "that more than 5,000 full sentients of the universal council participated in the discussion. They did not find your case to be without merit, but also recognized the harm that allowing such an inferior copy could do to the Earth brand. The deciding factor was the realization that this Earth has been so shoddily constructed that it will soon fall apart on its own. Plastic in the oceans, overpopulation, neoliberal economics … the planet is doomed, and any stay of destruction would therefore be moot."

"How long do we have?" I said.

"About a week, by your reckoning." He bowed his head. "I apologize for the inconvenience. Thank you for your time."

Then he and Floyd Standard turned around and left. ∎

**Timothy J. Gawne** *is a neuroscientist at the University of Alabama at Birmingham, the author of the Old Guy cybertank novels, and a Japanese Seiun Award nominee.*

ILLUSTRATION BY JACEY